Remarks

Claims 24, 26-28, 31-35 and 37-40 were rejected under 35 USC §112, first paragraph, as failing to comply with the written description requirement. Applicants respectfully submit that the recitation of "fungal disease resistance activity when expressed in a plant" does not constitute new matter.

It is stated on page 2, lines 5-19 of the specification that: "One such example is the Mlo gene of barley, which conveys resistance to *Erysiphe graminis f. sp. hordei*. The barley gene has been recently isolated by a positional cloning approach (Bueschges et al. (1997) *Cell 88*:695-705, copy enclosed). The dominant (sensitive) allele (Mlo) is thought to encode a protein involved in regulation of leaf cell death and in the onset of pathogen defense. The partial or complete inactivation of Mlo results in the priming of the disease-resistance response even in the absence of the pathogen, and leads to increased resistance to *E. graminis*."

Accordingly, since the specficiation discloses that the partial or complete inactivation of Mlo resulted in **disease resistance** of the Barley **plant** to the **fungus** *Erysiphe graminis* , it is respectfully submitted that this does constitute sufficient support for the recitation of "fungal disease resistance activity when expressed in a plant "

Withdrawal of the rejection of the claims under 35 USC paragraph 112, first paragraph, is respectfully requested in view of the above discussion

Claims 24, 26, 31-35 and 37-40 were rejected under 35 USC § 112, first paragraph, as failing to comply with the written description requirement.

It is respectfully submitted that the specification discloses to one of ordinary skill in the art a representative number of Mlo polypeptides having at least 90% sequence identity with the sequence set forth in SEQ ID NO:32, and not just a single polynucleotide encoding SEQ ID NO:32.

The specification, at page 7, line 3 to 32, discloses alterations in nucleotide sequence that are not expected to alter functionality, such as alterations that produce a chemically equivalent amino acid at a given site or alterations in the N- or C-terminal portions. Thus, from the foregoing, the skilled artisan would immediately understand the specification to disclose a representative number of polynucleotide sequences, having different nucleotide substitutions, that encode Mlo polypeptides but that vary (within 90% sequence identity) of SEQ ID NO:32.

Claims 24, 26-28, 31-35 and 37-40 were rejected under 35 USC §112, first paragraph, on the ground that the specification is not enabling for SEQ ID NO:32 and sequences recited 90-95% sequence identity.

Attention is kindly invited to Devoto *et al. JBC* (1999) 274: 34993-35004 (copy previously submitted) and Bueschges *et al.*, *Cell* (1997) 88:695-705 (copy

enclosed). The publications disclose seven distinctive Mlo transmembrane-spanning domains, a nuclear localization signal, which consists of short sequences containing mainly basic residues as described in Garcia *et al.* BBA (1991) 1071: 83-101 (copy enclosed for examiner's convenience) and two copies of the conserved casein kinase II motif (S/T-X-X-D/E) in addition to four Cysteine residues strictly conserved among all family members, located in extracellular loops 1 and 3 and indicative of their involvement in the formation disulfide bridges.

Enclosed herewith in Appendix A is a comparison of the claimed sequence with the barley sequence disclosed by Bueschges. This comparison demonstrates the sequence of the invention possesses these distinctive, highly conserved, Mlo regions.

Eleven mutagen induced Mlo resistant alleles were identified by Bueschges. The identified mutations were comprised of point mutations or deletions and all conferred pathogen resistance. The mutations clustered towards certain areas of the protein, which could be indicative of functionally sensitive domains. For the Examiner's convenience, all of the sites of mutation are labeled in the barley sequence in Appendix A. All the labeled residues are conserved among the barley and claimed sequence of SEQ ID NO:32. One skilled in the art would appreciate that the more highly conserved a residue is, the less likely that it could be modified and function maintained.

This alignment and the domains illustrate in the attached Appendix, that one of ordinary skill in the art could quickly determine which amino acid residues might be modified in SEQ ID NO:32 without a likely change in function. Since SEQ ID NOs:32 and the barley sequence share only 87% identity, one of skill in the art would have appreciated that many variants sharing at least 90% sequence identity to the SEQ ID NO:32 would have been expected to retain Mlo activity. Indeed, recently full complementation between barley and wheat Mlo mutants was achieved, confirming the expectation that Mlo from divergent species is functionally similar (Devoto et al. JMolEvol (2003) 56:77-88, copy enclosed for examiner's convenience). Thus, the known correlation of structure to function combined with the above discussion does enable one skilled in the art to make and use the claimed invention as commensurate in scope with the claims without undue experimentation.

It is respectfully submitted that the claims are now in form for allowance which allowance is respectfully requested.

Please charge any fees or credit any overpayment of fees which are required in connection with the filing of this Response Deposit Account No.: 04-1928 (E. I. du Pont de Nemours and Company).

Respectfully submitted,

LYNNE M. CHRISTENBURY
ATTORNEY FOR APPLICANTS
REGISTRATION NO.: 30,971
TELEPHONE NO.: 302-992-5481
FACSIMILE: 302-892-1026

Date: June 22, 2004

Cell, Vol. 88, 695–705, March 7, 1997

The Barley *Mlo* Gene: A Novel Control Element of Plant Pathogen Resistance

Rainer Büschges,*‖ Karin Hollricher,†‡‖
Ralph Panstruga,*‖ Guus Simons,‡‖
Marietta Wolter,† Adrie Frijters,‡
Raymond van Daelen,‡ Theo van der Lee,‡
Paul Diergaarde,‡ John Groenendijk,‡
Sonja Töpsch,† Pieter Vos,‡
Francesco Salamini,§ and Paul Schulze-Lefert*
*The Sainsbury Laboratory
John Innes Centre
Norwich Research Park Norwich NR4 7UH
United Kingdom
†Rheinisch-Westfälische Technische Hochschule
Institut für Biologie I
D-52074 Aachen
Germany
‡Keygene N.V.
P.O. Box 216
6700 AE Wageningen
Netherlands
§Max-Planck-Institut für Züchtungsforschung
Abt. für Züchtungsforschung und
Ertragsphysiologie
D-50829 Köln
Germany

Summary

Mutation-induced recessive alleles *(mlo)* of the barley *Mlo* locus confer a leaf lesion phenotype and broad spectrum resistance to the fungal pathogen, Erysiphe graminis f. sp. hordei. The gene has been isolated using a positional cloning approach. Analysis of 11 mutagen-induced *mlo* alleles revealed mutations leading in each case to alterations of the deduced Mlo wild-type amino acid sequence. Susceptible intragenic recombinants, isolated from *mlo* heteroallelic crosses, show restored *Mlo* wild-type sequences. The deduced 60 kDa protein is predicted to be membrane-anchored by at least six membrane-spanning helices. The findings are compatible with a dual negative control function of the Mlo protein in leaf cell death and in the onset of pathogen defense; absence of Mlo primes the responsiveness for the onset of multiple defense functions.

Introduction

In plants, resistance to specialized pathogens is frequently triggered by a recognition event followed by a coordinated complex defense response resulting in localized containment of the intruder (Hammond-Kosack and Jones, 1996). In this type of plant–pathogen interaction, resistance is specified by and dependent on the presence of two complementary genes, one from the host and one from the pathogen (Flor, 1971). The complementary genes have been termed race-specific resistance gene and avirulence gene, respectively. Several resistance genes have been isolated and appear to encode proteins that either contain a leucine-rich region (LRR), with or without an attached nucleotide binding site (NBS), indicative of ligand-binding and protein–protein interaction. Another class encodes a simple serine/threonine kinase (Dangl, 1995; Staskawicz et al., 1995; Zhou et al., 1995). The genetic and molecular observations are compatible with a specific receptor-mediated signal response triggering pathogen defense. The structural similarities in resistance gene products from different plant species to diverse pathogens such as bacteria, fungi, and viruses imply the existence of common "downstream" biochemical defense mechanisms. Although these mechanisms remain to be uncovered, localized death of host cells at the site of attempted infection, designated the hypersensitive response (HR), accompanies many incompatible race-specific interactions (Stakman, 1915; Staskawicz et al., 1995). Similarly, resistance in barley to the common biotrophic fungal pathogen Erysiphe graminis f. sp. hordei is in most analyzed cases specified by dominant or semidominant race-specific resistance genes and associated with a HR (*Mlx*; Jørgensen, 1994).

Monogenic resistance mediated by recessive (*mlo*) alleles of the *Mlo* locus is different. Apart from being recessive, it differs from race-specific incompatibility to single pathogen strains in that (1) it confers a broad spectrum resistance to almost all known isolates of the fungal pathogen, (2) *mlo* resistance alleles have been obtained by mutagen treatment of any tested susceptible wild-type (*Mlo*) variety, and (3) the resistance is apparently durable in the field despite extensive cultivation in Europe (Jørgensen, 1992). Finally, under pathogen-free or even axenic conditions, *mlo* plants exhibit a spontaneous leaf cell death phenotype, preceded by the appearance of characteristic cell wall appositions (Wolter et al., 1993).

Mutations have also been described in many other plant species in which cell death symptoms appear, resembling those in defense responses to plant pathogens (Walbot et al., 1983; Jones, 1994; Dangl et al., 1996). It has been suggested that at least some of these mutants, often collectively termed disease lesion mimics, affect control mechanisms of plant defense. Both recessively and dominantly inherited lesion mimic mutants have been analyzed for indicators of defense responses in Arabidopsis thaliana (Dietrich et al., 1994; Greenberg et al., 1994; Weymann et al. 1995). Apart from the onset of cell death in the absence of pathogens, multiple defense functions such as plant cell wall modifications and the accumulation of defense-related gene transcripts and phytoalexins have been observed. The mutants (*lsd1* to *lsd7* and *acd2*) were found to exhibit elevated resistance to a bacterial (Pseudomonas syringae) and a fungal (Peronospora parasitica) pathogen. Lesion mimic mutants are not restricted to foliar tissue. Recessive alleles of the soybean *Rn* locus exhibit, under axenic conditions, HR symptoms in the root, accompanied by the accumulation of defense-related proteins

‖These authors contributed equally to this work and are listed in an alphabetical order.

and the pythoalexin glyceollin (Kosslak et al., 1996). Homozygous *m* plants exhibit increased tolerance to root-borne infection by the fungal pathogen Phytophthora sojae. The findings suggest the activation of an at least partially overlapping set of biochemical events in pathogen-triggered race-specific resistance and during pathogen-independent cell death in several disease lesion mimic mutants.

We describe here the molecular isolation of the *Mlo* gene as a first step toward a molecular interpretation of broad spectrum resistance mediated by recessive host gene mutations. The gene encodes a member of a novel protein family apparently restricted to plants. We discuss the possible dual function of the Mlo protein in down-regulating leaf cell death and pathogen defense functions.

Results

We had previously identified RFLP markers closely linked to *Mlo* on barley chromosome 4 in *mlo* backcross (BC) lines containing *mlo* alleles from six genetic backgrounds (Hinze et al., 1991). The identification of a 2.7 cM (centiMorgan) RFLP interval (bAL88–bAO11) containing *Mlo* based on the cross Carlsberg II *Mlo* × Grannenlose Zweizeilige *mlo-11* opened a route to isolate the gene via positional cloning (Figure 1; RFLP map). However, because the barley genome has a very unfavorable ratio of genetic and physical distances (approximately 3 Mb/cM; Bennett and Smith, 1991; Becker et al., 1995), we applied AFLP (Amplified Fragment Length Polymorphism) marker technology (Vos et al., 1995) to increase the DNA marker density around *Mlo* and to generate a genetic map with a resolution better than 0.05 cM. We aimed to physically delimit the gene with flanking DNA markers on single large insert size genomic clones, an approach that has been termed "chromosome landing" (Tanksley et al., 1995).

Targeted Search for AFLP Markers

We selected AFLP markers around *Mlo* by searching for polymorphic DNA fragments between an *mlo* BC line (BC_7 Ingrid *mlo-3*) and DNA from the recurrent parent (Ingrid *Mlo*). The BC_7 Ingrid *mlo-3* line was previously shown to carry a small introgressed DNA segment on barley chromosome 4 (Hinze et al., 1991). The donor parent of the BC line represents a different genetic background (cultivar Malteria Heda *mlo-3*) in comparison to the recurrent parent line. In parallel, we established a second segregating F2 population from the cross Ingrid *Mlo* × BC_7 Ingrid *mlo-3*, formally representing an eighth backcross. To further narrow down the chromosomal interval for DNA marker identification to approximately 3 cM, pooled DNA from resistant (*mlo*) and susceptible (*Mlo*) F2 individuals were included in the search for AFLP markers besides DNA of the parental lines (see Experimental Procedures and Giovannoni et al., 1991). All possible PstI/MseI primer combinations (1,024) extending into genomic sequences up to nucleotide positions +2 and +3 and 880 EcoRI/MseI primer combinations (+3/+3) were tested. A total of 38 AFLP marker candi-

High Resolution Mapping

A three step procedure was chosen to construct the high resolution AFLP map. First, we were able to position 21 of the identified candidate AFLP markers to opposite sides of *Mlo* by using recombinants for flanking RFLP markers that had been detected among a small number of 70 F2 individuals (data not shown). The remaining 17 AFLP markers could not be separated from *Mlo* using this population size. In a second step, two codominant AFLP markers on opposite sides of *Mlo* were chosen to screen 2,022 F2 segregants for recombination events in the interval. 76 recombinants were identified, and their genotype at *Mlo* was determined by testing selfed F3 families with powdery mildew isolate K1 that is not virulent on homozygous *mlo* genotypes. In a third step, an AFLP analysis was carried out with each of the remaining 17 candidate AFLP markers to determine their position relative to *Mlo* based on the 76 recombinants. The crucial result was the identification of a DNA marker cosegregating with *Mlo* (Bpm16) and two flanking markers (Bpm2 and Bpm9) at a distance of 0.24 and 0.4 cM, respectively (Figure 1; AFLP map).

Physical Delimitation of *Mlo*

A large insert yeast artificial chromosome (YAC) library was constructed with genomic DNA of cultivar Ingrid *Mlo* using vector pYAC4 (Burke et al., 1987). The library comprises 40,000 clones with an average insert size of 500 kb and represents approximately four barley genome equivalents (construction and characterization of this library will be published elsewhere). Four YAC clones (YHV417-D1, YHV400-H11, YHV322-G2, and YHV303-A6) were isolated by an AFLP screen specific for marker Bpm16, which cosegregated with *Mlo*. AFLP analysis indicated that three of these clones (YHV400-H11, YHV322-G2, and YHV303-A6) also contained both flanking marker loci (Bpm2 and Bpm9). These findings implied physical delimitation of *Mlo* on three YAC clones.

We chose YHV303-A6 (insert size 650 kb; Figure 1) for subcloning experiments into bacterial artificial chromosome (BAC) vector pECSBAC4 containing a unique EcoRI site (Shizuya et al., 1992; see Experimental Procedures). Recombinant BAC clones containing the AFLP locus Bpm16 were subsequently identified using the cloned 108 bp PstI/MseI genomic Bpm16 fragment from cultivar Ingrid *Mlo* as a probe in colony hybridization experiments. One BAC clone, BAC F15, containing an insert of approximately 60 kb was chosen for further detailed studies (Figure 1; BAC F15). We found that the recombinant BAC clone contained locus Bpm2 in addition to the AFLP marker Bpm16, but not locus Bpm9, indicating physical delimitation in centromeric orientation to *Mlo*. Instead of constructing a BAC contig between Bpm16 and Bpm9, we developed new polymorphic markers from BAC F15 and mapped them using template DNA of 25 recombinants (derived from the high resolution mapping population described above) in the interval Bpm2–Bpm9. A codominant XbaI/MseI polymorphism (designated Bxm2) was identified between the parental lines Ingrid *Mlo* and BC_7 Ingrid *mlo-3*. The analysis of the 25 recombinant individuals revealed a position of Bxm2 in telomeric orientation from *Mlo* at a

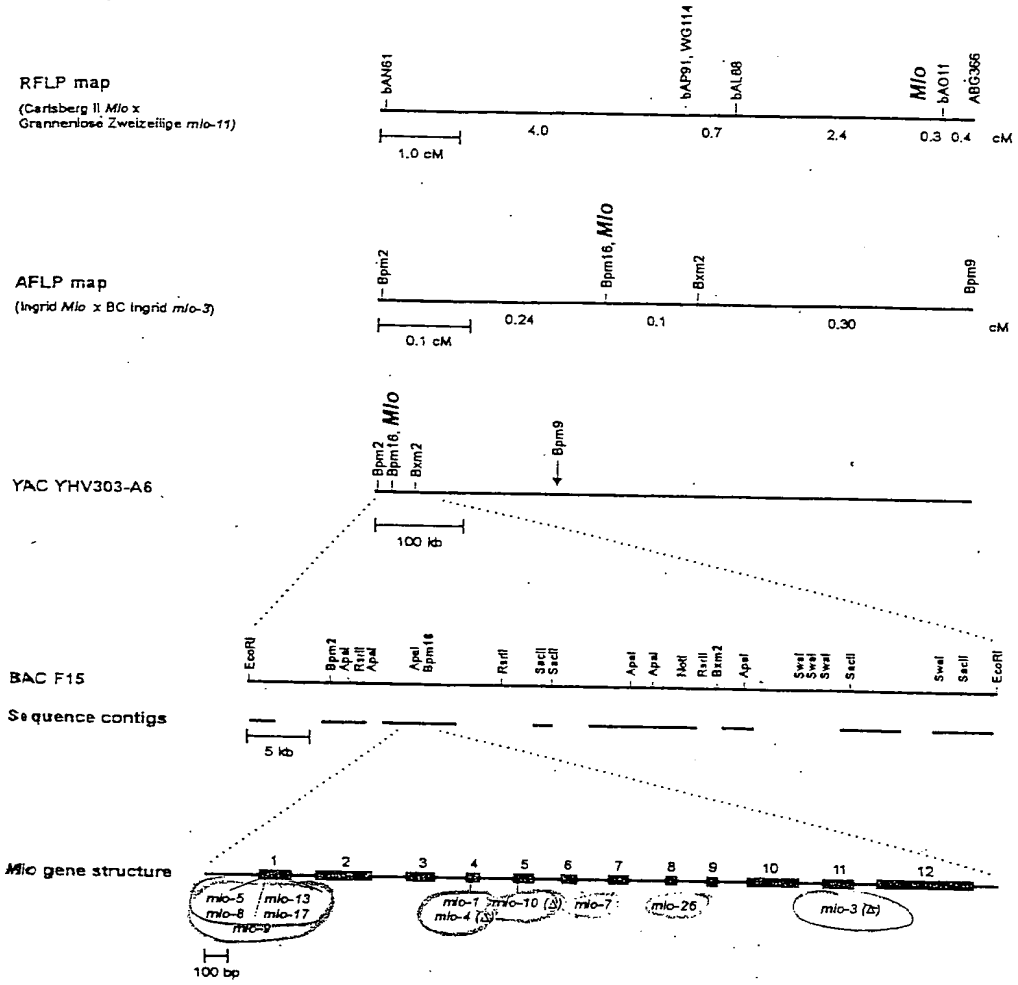

Figure 1. Positional Cloning of *Mlo*

The *Mlo* locus has been mapped with increasing precision on the long arm of barley chromosome 4 using RFLP and AFLP markers. The upper part of the figure presents the genetic linkage maps of these markers relative to *Mlo*. Genetic distances are indicated in centiMorgans (cM); the RFLP map is based on multipoint linkage analysis; and the AFLP map was calculated by two point estimates. The RFLP marker map is based on the analysis of 257 F2 individuals derived from the cross Carlsberg II *Mlo* × Grannenlose Zweizeilige *mlo-11*. The previously published RFLP map (see Results) of the same cross was based on only 44 F2 individuals. The gene was delimited to a 2.7 cM interval bordered by markers bAO11 (in telomeric orientation) and bAL88 (in centromeric orientation). AFLP markers (Bpm2, Bpm9, Bpm16, and Bxm2) were identified and mapped as described in Experimental Procedures. Their genetic distance to *Mlo* is based on the cross Ingrid *Mlo* × BC_7 Ingrid *mlo*-3. One YAC clone, YAC YHV303-A6 (insert size 650 kb), containing the cosegregating marker Bpm16 and two flanking loci (Bpm2 and Bpm9) is shown in the middle section of the figure. The position of marker Bpm9 was only roughly estimated within the YAC clone as indicated by the arrow. The insert of BAC F15 represents a 60 kb subfragment of this YAC as indicated in the lower part of the Figure. The approximate physical positions of AFLP markers Bpm2, Bpm16, and Bxm2 (spanning an interval of approximately 30 kb) as well as the location of some rarely occuring restriction sites are indicated. Dashed lines below the schematic representation of BAC F15 DNA show the position of the largest established DNA sequence contigs. The structure of the *Mlo* gene is given schematically in the bottom line of the figure. Exons are highlighted by closed boxes. Positions of mutational events are indicated for the eleven tested *mlo* alleles. Mutant alleles carrying deletions in their nucleotide sequences are marked with a (Δ).

distance of 0.1 cM (Figure 1; AFLP map). We concluded that *Mlo* had been physically delimited on BAC F15 between marker loci Bpm2 and Bxm2.

A Candidate *Mlo* Gene

DNA sequences of the approximately 60 kb insert of BAC F15 were obtained from randomly chosen clones of a plasmid sublibrary (see Experimental Procedures). In parallel, a physical map was generated (Figure 1; BAC F15). The map indicated that the flanking markers Bpm2 and Bxm2 are separated by approximately 30 kb. Rare cutting restriction sites enabled us to assign larger sequence contigs within BAC F15. We searched the available sequence contigs for regions of high coding probability (see Experimental Procedures). Only one sequence contig of 5.8 kb, including the cosegregating marker

Bpm16, revealed an extensive region of high coding probability.

We performed reverse transcriptase–polymerase chain reactions (RT–PCR) with total leaf RNA derived from cultivar Ingrid *Mlo* using a series of primers deduced from regions that indicated high coding probabilities and obtained in each case a distinct amplification product (Experimental Procedures). Sequencing of the largest RT–PCR products revealed a single extensive open reading frame of 1,599 bp (Figure 2). 5′ and 3′ ends of the gene transcript were identified using rapid amplification of cDNA ends (RACE) technology. The deduced putative protein of 533 amino acids has a molecular weight of 60.4 kDa. No significant homologies were found to any other described protein in the various databases, but at least six putative membrane-spanning helices indicated membrane association (for details, see Discussion). We were unable to detect a signal in Northern blot experiments containing total RNA with the labeled RT–PCR probe, but a rare RNA transcript of approximately 2.0 kb length was clearly visible in the tested *Mlo*, *mlo-1*, and *mlo-3* genotypes when poly(A)$^+$ RNA was used (Figure 3). This transcript size is in agreement with the combined data from RT–PCR and RACE analysis. A comparison of the genomic DNA and RT–PCR-derived sequences revealed 12 exons, each flanked by the consensus splice site sequences (Figures 1 and 2). Since marker Bpm16 is part of exon 11 and intron 11 and, as shown above, cosegregated with the resistance phenotype, it represented a candidate *Mlo* gene. We started genomic PCR-based sequencing of eleven mutagen-induced *mlo* resistance alleles and their corresponding wild-type DNAs (Experimental Procedures). These mutants had been isolated within six different genetic backgrounds. We identified nucleotide alterations (point mutations or deletions) in all tested mutant alleles that at the amino acid level result either in single amino acid substitutions or truncated versions of the predicted wild-type protein (Table 1). Surprisingly, a comparison among the wild-type gene sequences of seven tested barley cultivars (Carlsberg II, Diamant, Foma, Haisa, Ingrid, Malteria Heda, and Plena) indicated not a single amino acid difference. Moreover, we observed that at the nucleotide level the wild-type gene is identical among 6 tested cultivars both in exon and intron sequences, whereas cultivar Foma revealed 7 nucleotide substitutions (2 in exon and 5 in intron sequences). In conclusion, the comparative sequencing of genomic DNA from various mutant *mlo* lines and their respective *Mlo* wild-type cultivars supported our assumption that we had identified *Mlo*.

Characterization of Intragenic Recombinants

It had been our intention to provide a chain of evidence for the molecular isolation of *Mlo* that is not dependent upon complementation experiments by the time-consuming production of transgenic barley plants. We reasoned that recombination events between two physically separated mutation sites within the gene should give rise to a wild-type allele and an allele carrying both mutant sites. The former product of such rare intragenic recombination events is predicted to confer susceptibility upon powdery mildew attack only if the inactivation of the described candidate gene above is a requirement for resistance.

Based on this assumption, we performed intermutant crosses with lines containing alleles *mlo-1*, *mlo-5*, and *mlo-8*, generating in each case at least 10 F1 plants (Table 2; note that mutant sites in *mlo-1* and *mlo-5* as well as *mlo-1* and *mlo-8* are each separated by approximately 820 bp, as shown in Figure 1). The mutant alleles originate from the genetic backgrounds Haisa (*mlo-1*) and Carlsberg II (*mlo-5* and *mlo-8*). F2 populations were obtained by self-fertilization. F2 seedlings were screened for rare disease–susceptible individuals after inoculation with powdery mildew isolate K1, which is virulent on each of the parental *Mlo* wild-type cultivars (note that we were unable to select for products of intragenic recombination events carrying both mutagenic events because they are expected to exhibit a resistant phenotype). Susceptible F2 individuals were identified with an average frequency of 6×10^{-4}. This frequency is of the same order of magnitude as in previous reports of intragenic recombination events in plant genes (Salamini and Lorenzoni, 1970; Freeling, 1978; Koornneef et al., 1983; Dooner and Kermicle, 1986; Mourad et al., 1994). In contrast, when comparable numbers of progeny from selfings of each of the three *mlo* mutants were tested, no susceptible seedlings were identified (Table 2). This finding strongly indicated that the susceptible individuals derived from the intermutant crosses were not due to spontaneous reversion of the *mlo* alleles.

The inheritance of the susceptible F2 individuals was tested after selfing in F3 families. Each of the F2 individuals segregated in the F3 in the predicted ratio of 3 susceptible to 1 resistant, indicating heterozygosity for alleles conferring resistance and susceptibility in the F2. Homozygous susceptible F3 progeny were isolated for the majority of susceptible F2 individuals (see Experimental Procedures). A molecular analysis of these was performed using RFLP markers tightly linked (<4 cM) on each side of the *Mlo* locus to determine if restoration of *Mlo* function was accompanied by flanking molecular marker exchange (Figure 4). A compilation of the detected RFLP alleles of all relevant genotypes is given in Table 3. The compilation reveals that seven susceptible individuals exhibited flanking molecular marker exchange, indicating reciprocal crossover events (CO), whereas five susceptible individuals revealed no flanking marker exchange and therefore a non-crossover type of recombination (NCO). The latter class could be explained by a gene conversion or double crossover event. The ratio of the two observed classes (7:5) is compatible with the double-strand break repair model for recombination (Szostak et al., 1983). The relative position of the mutant sites in *mlo* alleles used in both heteroallelic crosses (Figure 1) predicts that the CO type recombination events are resolved unidirectionally with respect to flanking marker alleles in order to restore the *Mlo* wild-type allele. This is the case for all seven analyzed CO type recombinants (Table 3).

DNA of the CO type recombinants was tested for the presence of wild-type or mutant sequences. Genomic

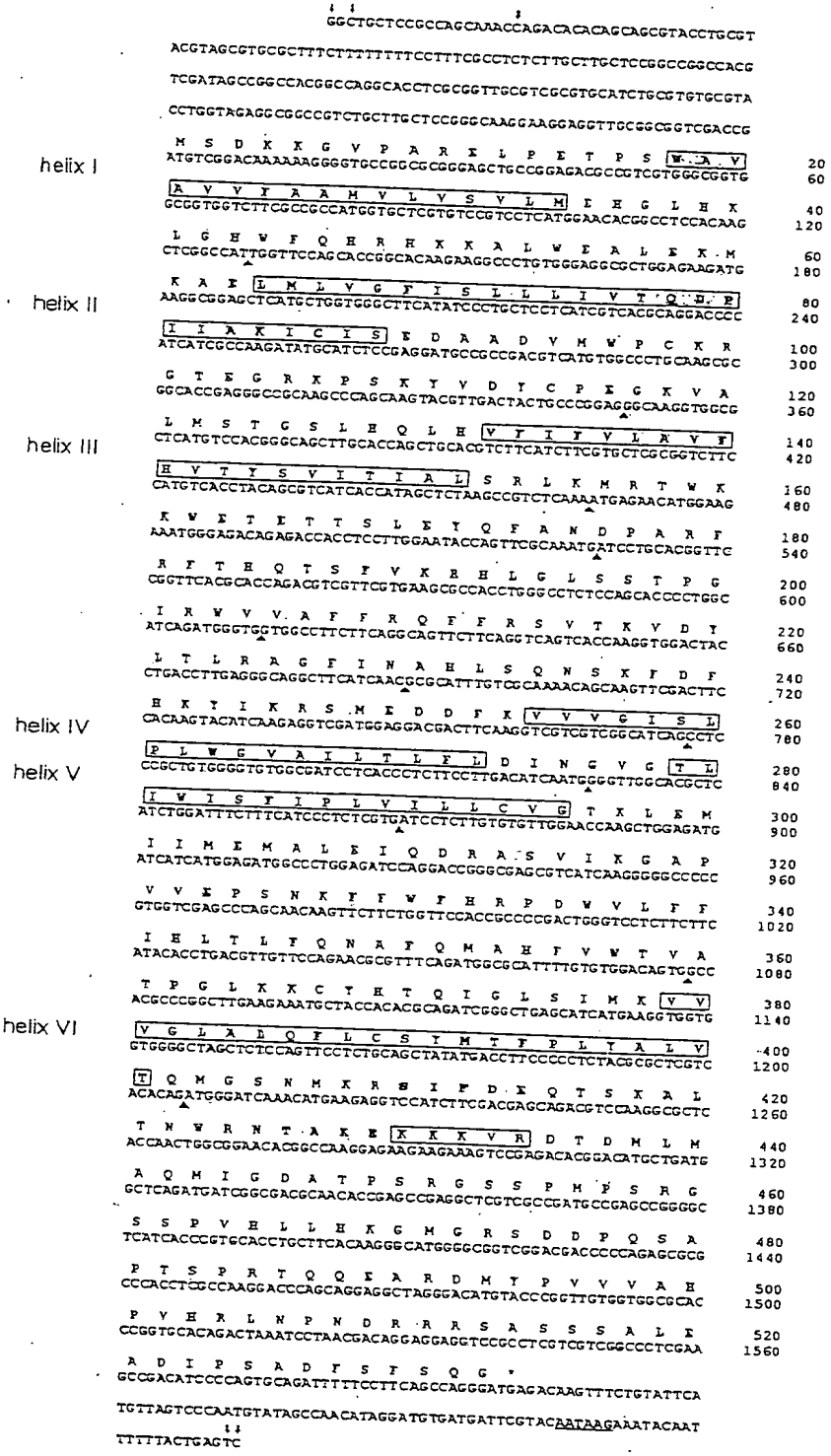

Figure 2. Nucleotide and Deduced Amino Acid Sequence of the Barley *Mlo* cDNA

The nucleotide and the deduced amino acid sequence are based on the combined data of RT–PCR and RACE obtained from experiments using RNA of cultivar Ingrid *Mlo*. The stop codon is marked by an asterisk, the putative polyadenylation signal is underlined, and the detected termini of RACE products are indicated by arrows above the sequence. Positions of introns as identified by comparison with corresponding genomic clones are labeled by triangles below the nucleic acid sequence. Six membrane-spanning helices predicted according to the MEMSAT algorithm (see Discussion) are boxed in gray. A putative nuclear localization signal (K-K-K-V-R) is boxed, and two casein kinase II sites (S/T-X-X-D/E) are shown in bold type.

PCR-based sequencing demonstrated in all cases restored wild-type sequences. This observation strongly suggested that the intragenic crossover event occured between nucleotides +1 and +821 in the cross *mlo-1* × *mlo-8* and between +3 and +821 in the cross *mlo-1* × *mlo-5* (numbers refer to genomic DNA sequences). Due

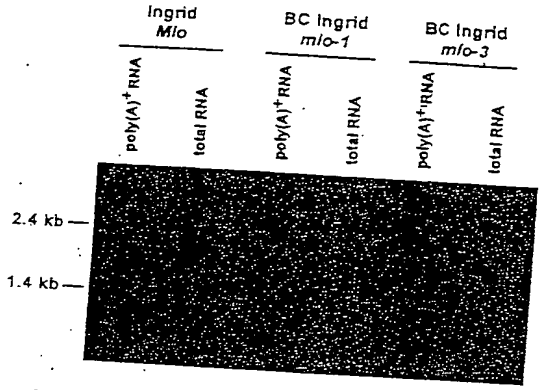

Figure 3. Northern Blot Analysis of *Mlo* Transcript Accumulation

Total RNA (20 μg) and poly(A)$^+$ RNA (5 μg) of 7 day old uninfected barley primary leaves of one wild-type (cultivar Ingrid *Mlo*) and two mutant (BC Ingrid *mlo-1*, BC Ingrid *mlo-3*) cultivars were isolated, separated on a 1.2 % formaldehyde gel, and transferred to a nitrocellulose membrane (Hybond). The filter was probed under stringent conditions with the radioactively labeled full size RT–PCR product derived from Ingrid *Mlo* (Figure 2). A clear signal is detected only in the lanes containing poly(A)$^+$ RNA. The signal corresponds to a size of approximately 2 kb.

to homomorphism of the DNA in these two intervals, we were unable to further delimit the intragenic recombination sites. In sum, the molecular analysis of seven intragenic recombinants from two heteroallelic crosses provides final proof that the above described candidate gene represents *Mlo*.

Discussion

We have described here an isolation procedure for the *Mlo* gene that is not dependent upon complementation experiments via transgenic barley plants. The chain of evidence rests on physical delimitation of *Mlo* to an approximately 30 kb interval by using flanking molecular markers and high resolution genetic mapping, identifying mutation sites in all tested *mlo* mutants, and demonstrating that susceptibility to the pathogen is coincident with restoration of the *Mlo* wild-type gene. Thus, the gene identification relies on reciprocal molecular tests involving both gene inactivation and restoration events and is conceptually similar to a transposon-tagging-mediated gene isolation approach based on gene inactivation by transposon insertion and restoration of gene function via transposon excision (Osborne and Baker, 1995).

The study has shown the feasibility of a chromosome landing approach (Tanksley et al., 1995) in the largest genome for which a positional cloning approach has been completed so far (5.3×10^9 bp/haploid genome equivalent, which is almost double the size of the human genome size; Bennett and Smith, 1991). Important elements were the construction of a local high resolution genetic map and the application of the AFLP marker technology (Vos et al., 1995), enabling us to delimit the target physically to 30 kb. The data reveal the variability of genome-wide and local ratios of genetic and physical distances. Within the *Mlo* gene, the recombination frequency was found to be 0.04 cM/kb (7 reciprocal crossover events among 20,670 F2 individuals in the 820 bp interval covering the first four exons, assuming that a comparable number of crossover events will have generated recombinant chromosomes carrying both mutant sites but were not detectable in the screen for susceptible intragenic recombinants). Thirteen recombinants were identified within the 30 kb interval bordered by markers Bxm2 and Bpm2 corresponding to 0.01 cM/kb, similar to the frequency found within the gene. These ratios deviate from the genome-wide estimate (0.0003 cM/kb) by one to two orders of magnitude (Bennett and Smith, 1991; Becker et al., 1995). Although the ratio of genetic/physical distances is generally believed to be much higher in telomeric regions of plant chromosomes (Heslop-Harrison, 1991), this argument does not apply to *Mlo* because of its location in the middle of the long arm of barley chromosome 4. Thus, the findings could be better explained with the reported exceptionally high frequencies of recombination within genes resulting in a scattered distribution of high and low recombination frequencies along a chromosome, as has been shown for the Arabidopsis thaliana chromosome 4 (Salamini and Lorenzoni, 1970; Freeling, 1978; Koornneef et al., 1983; Dooner and Kermicle, 1986; Schmidt et al., 1995).

The deduced amino acid sequence of *Mlo* reveals no homologies to any other described plant resistance gene so far, supporting the idea of a distinct mechanism triggering pathogen defense. Moreover, the gene shows no striking similarities to any characterized plant, animal, or prokaryotic gene in the various data bases (EMBL,

Table 1. *mlo* Mutant Alleles

Allele	Mother Variety	Mutagen	Mutational Event at *Mlo*	Effect on Amino Acid Level
mlo-1	Haisa	X-rays	$T^{484} \rightarrow A$	$Trp^{162} \rightarrow Arg$
mlo-3	Malteria Heda	γ-rays	Deletion of 2 nucleotides (1188–1189)	Frame shift after Phe^{395}
mlo-4	Foma	X-rays	Deletion of 11 nucleotides (478–488)	Frame shift after Trp^{159}
mlo-5	Carlsberg II	EMS	$G^{3} \rightarrow A$	$Met^{1} \rightarrow Ile^{a}$
mlo-7	Carlsberg II	EMS	$G^{677} \rightarrow A$	$Gly^{226} \rightarrow Asp$
mlo-8	Carlsberg II	EMS	$A^{1} \rightarrow G$	$Met^{1} \rightarrow Val^{a}$
mlo-9	Diamant	EMS	$C^{28} \rightarrow T$	$Arg^{10} \rightarrow Trp$
mlo-10	Foma	EMS	Deletion of 6 nucleotides (543–548)	2 amino acids (Phe^{182}, Thr^{183}) missing
mlo-13	Plena	γ-rays	$T^{89} \rightarrow A$	$Val^{30} \rightarrow Glu$
mlo-17	Plena	EMS	$C^{92} \rightarrow T$	$Ser^{31} \rightarrow Phe$
mlo-26	Plena	EMS	$T^{809} \rightarrow A$	$Leu^{270} \rightarrow His$

Numbers of nucleotides and amino acids are given according to the translational start site (see Figure 2). EMS = ethylmethane sulfonate.
[a] Next start codon is at nucleotide positions 79–81 and is in frame with the coding sequence.

Table 2. F2 Progeny Obtained from *mlo* Heterallelic Crosses and Corresponding *mlo* Selfings

Testcrosses[a]	Selfings	Resistant Individuals	Susceptible Individuals	Frequency of Susceptible F2 Progeny
mlo-8 × *mlo-1*		5,281	3	5.7×10^{-4}
mlo-1 × *mlo-5*		915	0	—
mlo-5 × *mlo-1*		14,474	9	6.2×10^{-4}
	mlo-1	12,634	0	—
	mlo-5	5,498	0	—
	mlo-8	8,435	0	—

[a] Crosses are given female × male.

GenBank, SWISS-PROT). However, highly significant homologous sequences have been identified both in the EST databases from rice and Arabidopsis thaliana (EMBL/GenBank accession numbers D24131, D24287, N37544, H76041, T22145, T22146, and T88073). In addition, we have isolated cross-hybridizing genomic clones from barley and rice containing highly homologous DNA sequences (data not shown). This strongly suggests that the Mlo protein is likely to represent a member of a novel protein family and implies a conserved function among monocot and dicot plants. We failed to detect homologous sequences in either the human, mouse, or Caenorhabditis EST data bases. Homologous sequences were also not detected in the Saccharomyces cerevisiae genome for which complete DNA sequence information is available (Dujon, 1996). Thus, Mlo is likely to represent a member of a novel protein family restricted to the plant kingdom.

A close inspection of the predicted amino acid sequence reveals six hydrophobic stretches that are likely to form at least six transmembrane helices (Figure 2). The significance of this finding is supported by applying three different algorithms for assessment of membrane-anchored proteins, indicating in each case six membrane-spanning helices (ALOM, Klein et al., 1985; MEMSAT, Jones et al., 1994; TMpred [http://ulrec3.unil.ch/software/TMPRED_form.html]). In addition, a putative nuclear localization motif (NLS) was found in exon 12, indicating a possible transport of the protein into the nucleus (K-K-K-V-R; Nigg et al., 1991). Two casein kinase II motifs (S/T-X-X-D/E; Rihs et al., 1991) are located immediately upstream of the NLS. Casein kinase II sites are frequently found at distances between 10 and 30 amino acids from NLS motifs and have been shown to determine the rate of nuclear transport (Rihs et al., 1991). However, because NLS motifs appear to be insufficient to target membrane-bound proteins to the nucleus (Soullam and Worman, 1995), detailed functional studies are necessary for subcellular localization of the protein.

The apparent clustering of the mutations in *Mlo* (Figure 1) may be the first hint that functionally sensitive domains in the protein can be delimited. The *mlo-4* allele, characterized by an 11 bp deletion in exon 4, might have a special implication. The resulting frameshift is predicted to shorten the length of the expressed Mlo

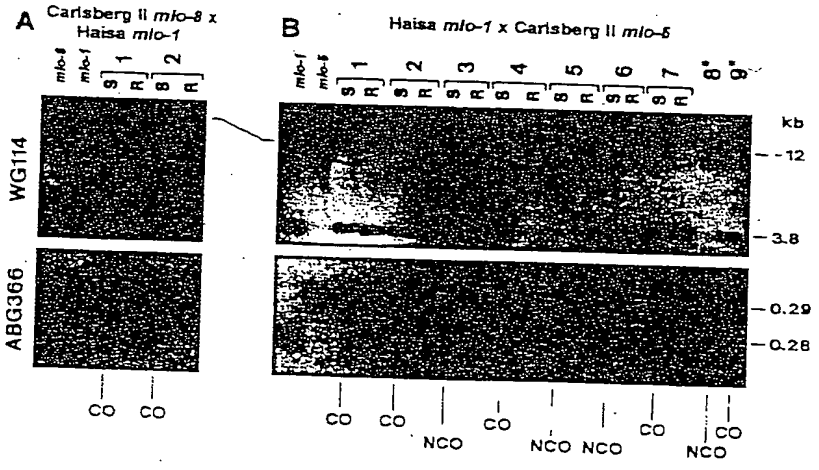

Figure 4. Southern Blot Analysis of Intragenic Recombinants Derived from *mlo* Heteroallelic Crosses

The alleles of two RFLP markers flanking *Mlo* on opposite sides of either susceptible F2 individuals or homozygous susceptible and homozygous resistant progeny were determined by Southern blot analysis. Plant DNA (10 μg) of the individuals were digested with PstI (A) or HaeIII (B) and hybridized with the radioactively labeled RFLP markers WG114 (upper panel; maps 3.1 cM in centromeric orientation to *Mlo*; see Figure 1) and ABG366 (lower panel; maps 0.7 cM in telomeric orientation to *Mlo*; see Figure 1) according to standard procedures. CO, crossover type of recombinants; NCO non-crossover type of recombinants.

(A) DNA of the parental lines *mlo-8* and *mlo-1* as well as 2 homozygous susceptible (S, *Mlo Mlo*) and 2 resistant (R, *mlo mlo*) progenies derived from 2 susceptible F2 plants (designated 1 and 2) were tested. The DNAs in lanes S and R represent selected F3 individuals from F3 families obtained by selfing the susceptible F2 individuals 1 and 2. Note that susceptible F2 individuals are expected to be heterozygous at *Mlo* in this selection scheme. Infection phenotypes were scored 7 days after inoculation with the *mlo* avirulent isolate K1. DNA from a third susceptible individual of this heteroallelic cross (see Table 3) is not included in this Figure.

(B) DNA of the parental lines *mlo-5* and *mlo-1* and 7 homozygous susceptible (S, *Mlo Mlo*) and 7 resistant (R, *mlo mlo*) progeny derived from 7 susceptible F2 plants (designated 1 to 7) were tested. The DNAs in lanes S and R represent selected F3 individuals from F3 families obtained by selfing the susceptible F2 individuals 1–7. DNA was analyzed from 2 further susceptible individuals of this heteroallelic cross only in the F2 generation (8* and 9*).

Table 3. Genotypes at Flanking RFLP Markers in Susceptible Progeny Derived from Heteroallelic *mlo* Crosses

Testcrosses[a]	Susceptible Plants	Parental Genotype in Centromeric Orientation to *Mlo*[b]	Parental Genotype in Telomeric Orientation to *Mlo*[c]	Type of Recombination
mlo-8 × *mlo-1*	1	*mlo-1*	*mlo-8*	CO
	2	*mlo-1*	*mlo-8*	CO
	3	*mlo-8*	*mlo-8*	NCO
mlo-1 × *mlo-5*	1	*mlo-1*	*mlo-5*	CO
	2	*mlo-1*	*mlo-5*	CO
	3	*mlo-5*	*mlo-5*	NCO
	4	*mlo-1*	*mlo-5*	CO
	5	*mlo-5*	*mlo-5*	NCO
	6	*mlo-5*	*mlo-5*	NCO
	7	*mlo-1*	*mlo-5*	CO
	8[d]	*mlo-5*	*mlo-5*	NCO
	9[d]	*mlo-1* + *mlo-5*	*mlo-5*	CO

CO = crossover type, NCO = noncrossover type of recombination.
[a] Crosses are given female × male.
[b] Deduced from alleles of RFLP marker WG114 (see Figure 1).
[c] Deduced from alleles of RFLP marker ABG366 (see Figure 1).
[d] Genotypes of flanking RFLP markers have been determined in heterozygous susceptible F2 individuals; in all other cases, homozygous susceptible F3 progeny derived from the susceptible F2 individuals were tested.

protein by 75%. We assume that at least this resistance allele represents a complete functional inactivation of the protein.

This study has shown that broad spectrum resistance to the powdery mildew fungus is caused by a defective *Mlo* gene. One of the key questions concerns function(s) of the Mlo protein and its homologues. Based on experimental evidence, we propose two hypotheses. In the first model, Mlo would have a negative control function in leaf cell death since punctate dead cell leaf lesions appear even in axenically grown seedlings carrying different *mlo* alleles (Wolter et al., 1993). In analogy to pcd in animals (Raff, 1992; White, 1996), Mlo would suppress a default cell suicide program in foliar tissue. In this scenario, resistance would have to be envisaged as a consequence of deregulated pcd. The intimate link between resistance and extent of spontaneous leaf lesions has been studied on the basis of an allelic series of 95 chemically induced *mlo* alleles (Habekuss and Hentrich, 1988). The defective alleles could be classified according to gradually different infection phenotypes upon infection of a mixture of nine powdery mildew isolates. Only three mutant alleles were found to exhibit an intermediate infection phenotype (i.e., a considerable number of sporulating fungal colonies on the leaf surface) and revealed no macroscopically detectable leaf lesions. In contrast, the most efficient resistance alleles exhibited pronounced necrosis. Another line of evidence for a connection between resistance and deregulated cell death control in *mlo* mutants comes from recently identified genes (*Ror1* and *Ror2*) that are required for *mlo* function, i.e., resistance to the pathogenic fungus (Freialdenhoven et al., 1996). The dead cell leaf lesion phenotype in pathogen-free grown *Ror mlo* mutants is abolished in *ror mlo* double mutants as determined by a failure for trypan blue uptake in the latter genotype (P. S.-L. et al., unpublished data).

In our second model, the Mlo protein would have a specific negative regulatory function by down-regulating multiple defense-related functions. Spontaneous cell death in *mlo* genotypes would merely represent the fatal end of an accumulating activation of defense responses. It is supported by the chronological order of defense-related events in *mlo* genotypes in the absence of pathogens. Cell wall appositions (CWAs) appear spontaneously in epidermal tissue of approximately 14 day old seedlings (Wolter et al., 1993). CWAs are always found in response to an authentic penetration attempt of the pathogen directly beneath the fungal appressorium and are believed to form a physical barrier against pathogen ingress (Bayles et al., 1990). In 18 day old seedlings, trypan blue positive leaf cell patches appear, indicating commitment to cell death, and 2–3 days later, necrotic flecks become macroscopically detectable (P. S.-L. et al., unpublished data). However, powdery mildew resistance is fully functional in 5 day old *mlo* seedlings, the earliest time point to carry out a resistance test for technical reasons. Thus, establishment of defense-associated events in pathogen-free *mlo* genotypes is not a requirement for effective resistance upon attempted powdery mildew attack. We conclude that a complete or partial inactivation of the Mlo protein "primes" or up-regulates the responsiveness of the seedling for the onset of pathogen defense.

This is an important difference to all but one characterized lesion mimic in Arabidopsis as well as to the soybean *rm* mutants, in which elevated resistance to pathogens is either dependent on lesion formation (*acd2*, *lsd2*, *lsd3*, *lsd4*, *lsd5*, and *rm*; Greenberg et al., 1994; Dietrich et al., 1994; Kosslak et al., 1996) or is expressed concomitantly with the appearance of dead cell lesions (*lsd6* and *lsd7*; Weymann et al. 1995). So far, only the Arabidopsis *lsd1* lesion mimic mutant appears to exhibit elevated pathogen resistance at the prelesion state (Dietrich et al., 1994). In contrast to the determinate and punctate growth of lesions in *mlo* leaves, lesion formation is indeterminate in *lsd1*, consuming the entire leaf. However, host cell death cannot be required for the early developmental arrest of the powdery mildew fungus in CWAs in the genuine interaction or *mlo*

the host cell survives the attack (Jørgensen and Mortensen, 1977; Wolter et al., 1993). A priming of defense functions in *mlo* plants would make it possible that inefficient defense responses in the *Mlo* genotype (e.g., CWA formation) become efficient (e.g., through increased speed and/or reduced response times) but that the early developmental arrest of the pathogen is insufficient to trigger a signal for the execution of the host cell death reaction.

Both of these two seemingly different functions of the Mlo protein could be explained by assuming that the protein has a dual function in down-regulating onset of leaf cell death and onset of multiple defense functions. We expect that the biochemical characterization of Mlo, the analysis of the unique collection of *mlo* mutants, and the identification of proteins with which it physically interacts should provide further insight into the molecular relationship between the control of plant cell death and plant defense functions.

Experimental Procedures

Plant Material

A compilation of the *mlo* mutants and their mother varieties analyzed in this study has been described by Jørgensen (1992) (*mlo-1*, *mlo-3*, *mlo-4*, *mlo-5*, *mlo-7*, *mlo-8*, *mlo-9*, *mlo-10*, and *mlo-11*) and by Habekuss and Hentrich (1988) (mutants in cultivar Plena 2018 [*mlo-13*], 2034 [*mlo-17*], 2118). Since mutant 2118 has not been assigned to an allele number so far, we designate the allele here as *mlo-26*, according to current numbering in the GrainGene database (gopher://greengenes.cit.cornell.edu). All *mlo* BC lines in cultivar Ingrid were a gift from Prof. James McKey, Uppsala, Sweden.

The high resolution map is based on a cross between Ingrid *Mlo* × BC_7 Ingrid *mlo-3* . F1 plants were selfed generating a segregating F2 population of approximately 600 plants. Phenotypically susceptible F2 plants that showed heterozygosity for RFLP markers on opposite sites of *Mlo* were selfed and generated further segregants in the F3 generation for high resolution mapping.

Powdery Mildew Infection Tests

The fungal isolate K1 (Hinze et al., 1991) is virulent on all cultivars used in this study carrying the *Mlo* allele and avirulent on all tested *mlo* genotypes. Plant growth and inoculation with Erysiphe graminis f. sp. hordei were carried out as described previously (Freialdenhoven et al., 1996). The genotype at *Mlo* of recombinants used for the high resolution map were determined after selfing and subsequent inoculation experiments in F3 or F4 families comprising at least 24 individuals.

AFLP Analysis

Genomic DNA for AFLP analysis was isolated according to Stewart and Via (1993). AFLP analysis was carried out as described by Vos et al. (1995). A set of four DNA templates has been used: from the susceptible parent cultivar Ingrid *Mlo*, the resistant parent BC_7 Ingrid *mlo-3*, a pool of 2 resistant F2 individuals (*mlo-3 mlo-3*), and a pool of 9 susceptible F2 individuals (*Mlo Mlo*) derived from the cross Ingrid *Mlo* x BC_7 Ingrid *mlo-3* . Amplified genomic fragments representing AFLP markers Bpm2, Bpm9, and Bpm16 (Figure 1) were cloned and sequenced as follows: gel pieces (fixed by vacuum drying to Whatman 3MM paper) containing the amplified genomic fragments were identified via autoradiography and subsequently excised. 100 μl water were added and boiled for 10 min, and after centrifugation, 5 μl of the supernatant was used as a template for nonradioactive reamplification (30 cycles) with the selective AFLP primers. Amplification products were isolated after agarose gel electrophoresis and subsequently cloned in the EcoRV site of pBluescript SK (Stratagene). Sequencing reactions were performed using a dye terminator cycle sequencing reaction kit (Perkin-Elmer) and resolved either on an ABI 373 or 377 (Applied Biosystems) automated sequencer.

Barley YAC Library and BAC Sublibrary Construction of YAC YHV303-A6

The YAC library of barley cultivar Ingrid was established using the pYAC4 vector (Burke et al., 1987) and yeast strain AB 1380. Details of the library construction and its characterization will be described elsewhere. Screening for YAC clones containing marker Bpm16 was done by AFLP analysis. For construction of a BAC sublibrary of YAC YHV303-A6, total DNA of this yeast clone was used. After partial EcoRI digestion and preparative pulsed-field gel electrophoresis, DNA fragments in the size range of 50 kb were recovered and subcloned in the pECSBAC4 vector. Clones carrying YHV303-A6-derived inserts were identified by a two step colony hybridization procedure using first, labeled total DNA of the nonrecombinant yeast strain AB 1380 and subsequently, labeled recombinant chromosome YHV303-A6 after enrichment by preparative pulsed-field gel electrophoresis.

DNA Sequencing of BAC F15

DNA of BAC F15 was isolated by an alkaline lysis large scale plasmid preparation according to Sambrook et al. (1989). 50 μg of purified DNA were nebulized by high pressure treatment with argon gas in a reaction chamber for 150 s. The ends of the sheared and reprecipitated DNA were blunt-ended by a T4 DNA polymerase–mediated fill-in reaction. DNA fragments in the size range of 800 bp to 3 kb were isolated from agarose gels using a DNA isolation kit (Jetsorb, Genomed Inc., USA), subcloned into the pBluescript SK vector (Stratagene), and propagated in E. coli DH5α. Clones carrying BAC F15–derived inserts were selected by hybridization using the labeled sheared DNA of BAC F15 as a probe. Sequencing reactions were performed as described above. Evaluation of the sequencing data, construction of sequence contigs, and estimation of coding propabilities were done by means of the STADEN software package for Unix users (fourth edition, 1994). Homology searches were done using the BLAST software.

PCR-Based Sequencing of Alleles at *Mlo*

Plant chromosomal DNA for this purpose was isolated according to Chunwongse et al. (1993). DNA sequences of *Mlo* alleles of the different barley varieties, *mlo* mutants, BC lines, and intragenic recombinants used in this study were obtained by PCR-based sequencing. Using sets of specific primers, seven overlapping subfragments of the gene (each 400–600 bp in length) were amplified by PCR (35 cycles, 60°C annealing temperature). After preparative agarose gel electrophoresis and isolation of the amplification products using the Jetsorb kit (Genomed Inc., USA), fragments were reamplified. The resulting products were subsequently purified from nucleotides and oligonucleotides (Jetpure, Genomed Inc., USA) and used as a template in DNA sequencing reactions (see above). All DNA sequences of mutant alleles and corresponding regions of the parental lines and the intragenic recombinants were derived from both strands and confirmed in independent sets of experiments.

RT–PCR and Rapid Amplification of cDNA Ends (RACE)

RT–PCR was performed using the SUPERSCRIPT preamplification system for first-strand cDNA synthesis (Gibco BRL). Total RNA (1 μg) of 7 day old primary barley leaves (cultivar Ingrid) served as template. First-strand cDNA synthesis was primed by an oligo(dT) primer. The putative coding region of the *Mlo* gene was subsequently amplified using oligonucleotides 25L (GTGCATCTGCGTGTGCGTA) and 38 (CAGAAACTTGTCTCATCCCTG) in a single amplification step (35 cycles, 60°C annealing temperature). The resulting product was analyzed by direct sequencing. 5′ and 3′ ends of the *Mlo* cDNA were determined by RACE using the MARATHON cDNA amplification kit (Clontech). To obtain specific RACE products, two consecutive rounds of amplification (35 cycles, 55°C annealing temperature) were necessary. Two sets of nested primers were used in combination with the adapter primers of the kit: oligonucleotides 46 (AGGGTCAG GATCGCCAC) and 55 (TTGTGGAGGCCGTGTTCC) for the 5′ end and primers 33 (TGCAGCTATATGACCTTCCCCCTC) and 37 (GGA CATGCTGATGGCTCAGA) for the 3′ end. RACE products were subcloned into pBluescript SK (Stratagene). Ten 5′ end and eight 3′ end clones were chosen for DNA sequence analysis.

Acknowledgments

Correspondence regarding this paper should be addressed to P. S.-L. We thank our colleagues Colwyn Thomas, Jonathan Jones, and Martin Parniske at the Sainsbury Laboratory for helpful suggestions on the manuscript. Leony Chamberlain, Paddy Bovill, and David Baker are acknowledged for their excellent technical assistance. We are grateful to James McKey for providing us with the valuable *mlo* backcross line material. This work was supported by grants from the Deutsche Forschungsgemeinschaft Schwerpunktprogramm "Molekulare Phytopathologie" and the Gatsby Foundation to P. S.-L.

Received November 14, 1996; revised January 16, 1997.

References

Bayles, C.J., Ghemawat, M.S., and Aist, J.R. (1990). Inhibition by 2-deoxy-D-glucose of callose formation, papilla deposition, and resistance to powdery mildew in *ml-o* barley mutant. Physiol. Mol. Plant Pathol. *36*, 63–72.

Becker, J., Vos, P., Kuiper, M., Salamini, F., and Heun, M. (1995). Combined mapping of AFLP and RFLP markers in barley. Mol. Gen. Genet. *249*, 65–73.

Bennett, M.D., and Smith, J.B. (1991). Nuclear DNA amounts in angiosperms. Phil. Trans. Roy. Soc. (Lond.) B *334*, 309–345.

Burke, D.T., Carle, G.F., and Olson, M.V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science *236*, 806–812.

Chunwongse, J., Martin, G.B., and Tanksley, S.D. (1993). Pre-germination genotypic screening using PCR amplification of half seeds. Theor. Appl. Genet. *86*, 694–698.

Dangl, J.L., Dietrich, R.A., and Richberg, M.H. (1996). Death don't have no mercy: cell death programs in plant–microbe interactions. Plant Cell *8*, 1793–1807.

Dangl, J.L. (1995). Piéce de rèsistance: novel classes of plant disease resistance genes. Cell *80*, 363–366.

Dietrich, R.A., Delaney, T.P., Ukness, S.J., Ward, E.R., Ryals, J.A., and Dangl, J.L. (1994). Arabidopsis mutants simulating disease resistance response. Cell *77*, 565–577.

Dooner, H.K., and Kermicle, J.L. (1986). The transposable element *Ds* affects the pattern of intragenic recombination at the *BZ* and *R* loci in maize. Genetics *113*, 135–143.

Dujon, B. (1996). The yeast genome project: what did we learn? Trends Genet. *12*, 263–270.

Flor, H.H. (1971). Current status of the gene-for-gene concept. Annu. Rev. Phytopathol. *9*, 275–296.

Freeling, M. (1978). Allelic variation at the level of intragenic recombination. Genetics *89*, 211–224.

Freialdenhoven, A., Peterhänsel, C., Kurth, J., Kreuzaler, F., and Schulze-Lefert, P. (1996). Identification of genes required for the function of non-race-specific *mlo* resistance to powdery mildew in barley. Plant Cell *8*, 5–14.

Giovannoni, J.J., Wing, R.A., Ganal, M.W., and Tanksley, S.D. (1991). Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. Nucl. Acids Res. *19*, 6553–6558.

Greenberg, J.T., Guo, A., Klessig, D.F., and Ausubel, F.M. (1994). Programmed cell death in plants: a pathogen-triggered response activated coordinately with multiple defense functions. Cell *77*, 551–563.

Habekuss, A., and Hentrich, W. (1988). Charakterisierung funktionell verschiedener ml-o Mutanten durch Primärinfektion, Pustelwachstum, Inkubationszeit und Befallsverlauf. Tag.-Ber., Akad. Landwirtsch.-Wiss. DDR, Berlin *272*, 229–237.

Hammond-Kosack, K.E., and Jones, J.D.G. (1996). Resistance gene–dependent plant defense responses. Plant Cell *8*, 1773–1791.

Heslop-Harrison, J.S. (1991). The molecular cytogenetics of plants. J. Cell Sci. *100*, 15–21.

Hinze, K., Thompson, R.D., Ritter, E., Salamini, F., and Schulze-Lefert, P. (1991). RFLP-mediated targeting of the *mlo* resistance locus in barley (*Hordeum vulgare*). Proc. Natl. Acad. Sci. USA *88*, 3691–3695.

Jones, J.D.G. (1994). Paranoid plants have their genes examined. Curr. Biol. *4*, 749–751.

Jones, D.T., Taylor, W.R., and Thornton, J.M. (1994). A model recognition approach of all-helical membrane protein structure and topology. Biochemistry *33*, 3038–3049.

Jørgensen, J.H., and Mortensen, K. (1977). Primary infection by *Erysiphe graminis* f. sp. *hordei* of barley mutants with resistance genes in the *mlo* locus. Phytopathology *67*, 678–685.

Jørgensen, J.H. (1992). Discovery, characterization and exploitation of *Mlo* powdery mildew resistance in barley. Euphytica *63*, 141–152.

Jørgensen, J.H. (1994). Genetics of powdery mildew resistance in barley. Crit. Rev. Plant Sci. *13*, 97–119.

Klein, P., Kanehisa, M., and DeLisi, C. (1985). The detection and classification of membrane-spanning proteins. Biochim. Biophys. Acta *815*, 468–476.

Koornneef, M., Van Eden, J., Hanhart, C.J., and De Jongh, A.M.M. (1983). Genetic fine-structure of the *GA-1* locus in the higher plant *Arabidopsis thaliana* (L.) Heynh. Genet. Res. Camb. *41*, 57–68.

Kosslak, R.M., Dieter, J.R., Ruff, R.L., Chamberlin, M.A., Bowen, B.A., and Palmer, R.G. (1996). Partial resistance to root-borne infection by *Phytophthora sojae* in three allelic necrotic root mutants in soybean. J. Hered. *87*, 415–422.

Mourad, G., Haughn, G., and King, J. (1994). Intragenic recombination of the *CSR1* locus of *Arabidopsis*. Mol. Gen. Genet. *243*, 178–184.

Nigg, E.A., Baeuerle, P.A., and Luhrmann, R. (1991). Nuclear import–export: in search of signals and mechanisms. Cell *66*, 15–22.

Osborne, B.I., and Baker, B. (1995). Movers and shakers: maize transposons as tools for analyzing other plant genomes. Curr. Opin. Cell Biol. *7*, 406–413.

Raff, M.C. (1992). Social controls on cell survival and cell death. Nature *356*, 397–400.

Riihs, H.-P., Jans, D.A., Fan, H., and Peters, R. (1991). The rate of nuclear cytoplasmic protein transport is determined by the casein kinase II site flanking the nuclear localization sequence of the SV40 T-antigen. EMBO J. *10*, 633–639.

Salamini, F., and Lorenzoni, C. (1970). Genetical analysis of glossy mutants of maize. Mol. Gen. Genet. *108*, 225–232.

Sambrook, J., Fritsch, E.F., and Manniatis, T. (1989). Molecular Cloning: A Laboratory Manual, Second Edition (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Schmidt, R., West, J., Love, K., Lenehan, Z., Lister, C., Thompson, H., Bouchez, D., and Dean, C. (1995). Physical map and organization of *Arabidopsis thaliana* chromosome 4. Science *270*, 480–483.

Shizuya, H., Birren, B., Kim, U.-J., Mancino, V., and Slepak, T. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc. Natl. Acad. Sci. USA *89*, 8794–8797.

Soullam, B., and Worman, H.J. (1995). Signals and structural features involved in integral membrane protein targeting to the inner nuclear membrane. J. Cell Biol. *130*, 15–27

Stakman, E.C. (1915). Relation between *Puccinia graminis* and plants highly resistant to its attack. J. Agric. Res. *4*, 193–199.

Staskawicz, B.J., Ausubel, F.M., Baker, B.J., Ellis, J.G., and Jones, J.D.G. (1995). Molecular genetics of plant disease resistance. Science *268*, 661–667.

Stewart, C.N., and Via, L.E. (1993). A rapid CTAB DNA isolation technique for RAPD fingerprinting and other PCR applications: BioTechniques *14*, 748–750.

Szostak, J.W., Orr-Weaver, T.L., and Rothstein, R.J. (1983). The double-strand-break repair model for recombination. Cell *33*, 25–35.

Tanksley, S.D., Ganal, M.W., and Martin, G.B. (1995). Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. Trends Genet. *11*, 63–68.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Homes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. Nucl. Acids Res. *23*, 4407–4414.

Walbot, V., Hosington, D.A., and Neuffer, M.G. (1983). Disease lesion mimic mutants. In Genetic Engineering of Plants, T. Kosuge, C.P. Meredith, and A. Hollaender, eds. (New York: Plenum Press), pp. 431–442.

Weymann, K., Hunt, M., Uknes, S., Neuenschwander, U., Lawton, K., Steiner, H.-Y., and Ryals, J. (1995). Suppression and restoration of lesion formation in *Arabidopsis lsd* mutants. Plant Cell *7*, 2013–2022.

White, E. (1996). Life, death, and the pursuit of apoptosis. Genes Dev. *10*, 1–15.

Wolter, M., Hollricher, K., Salamini, F., and Schulze-Lefert, P. (1993). The *mlo* resistance alleles to powdery mildew infection in barley trigger a developmentally controlled defence mimic phenotype. Mol. Gen. Genet. *239*, 122–128.

Zhou, J., Loh, Y.-T., Bressan, R.A., and Martin, G.B. (1995). The tomato gene *Pti1* encodes a serine/threonine kinase that is phosphorylated by Pto and is involved in the hypersensitive response. Cell *83*, 925–935.

EMBL/GenBank Accession Number

The accession number for the sequence of *Mlo* is Z83834.

Biochimica et Biophysica Acta, 1071 (1991) 83–101

ADONIS 0304415791000544

BBAREV 85077

Review

Nuclear protein localization

Jose Garcia-Bustos, Joseph Heitman and Michael N. Hall

Department of Biochemistry, Biocenter, University of Basel, Basel (Switzerland)

(Received 8 May 1990)

Contents

Introduction

Nuclear protein localization is a relatively new topic in the general field of protein transport. Despite the late start, many significant advances have already been made, and the study of nuclear protein transport is now precociously making the transition from an immature stage of observation and definition to a more mature stage in which meaningful statements about molecular mechanisms can be made. This rapid maturation can be attributed to a wide-spread interest in nuclear protein localization and the fortunate circumstance that the site of nuclear import, the nuclear pore complex, is unambiguous. Here we review the experimental approaches, findings and conclusions which have shaped the current understanding of nuclear protein localization. The reader is referred to other reviews for additional infor-

Abbreviations: BSA, bovine serum albumin; HSA, human serum albumin; HSP, heat shock protein; IIF, indirect immunofluorescence; NEM, *N*-ethylmaleimide; NLS, nuclear localization signal; NPC, nuclear pore complex; WGA, wheat germ agglutinin.

Correspondence: M.N. Hall, Dept. Biochemistry, Biocenter, University of Basel, CH-4056 Basel, Switzerland.

mation and perspectives on nuclear import [11,31,55, 125,137], the nuclear envelope [45,49,67,98] or the nucleus in general [68,109].

Nuclear proteins are synthesized in the cytoplasm and posttranslationally transported across the nuclear envelope which separates the cytoplasm from the nucleoplasm [31,41]. The nuclear envelope consists of two concentric membranes (the inner and outer membranes), nuclear pore complexes and the nuclear lamina [49]. The inner and outer membranes, which appear by electron microscopy as conventional lipid bilayers, are separated by a so-called perinuclear space which is continuous with the lumen of the endoplasmic reticulum (ER). The outer membrane is continuous with the ER and is functionally similar. The nuclear pore complex is a large proteinaceous, grommet-like structure that traverses the nuclear envelope at sites where the inner and outer membranes are fused, thereby creating an aqueous channel through which proteins (and RNA) cross the nuclear envelope. Transport into the nucleus is thus most likely fundamentally different from the transport of proteins into other organelles, or to the cell exterior, in which transported proteins presumably pass directly through the hydrophobic environment of a membrane [148]; a large, transport-mediating pore structure such as the nuclear pore complex is unique to the nuclear envelope. The nuclear lamina is a filamentous protein meshwork which lines the nucleoplasmic surface of the inner membrane. Both the inner membrane and the nuclear pore complex are anchored to the lamina.

Transport across the nuclear envelope is an active process mediated by a nuclear localization signal (NLS) contained within the transported protein. Nuclear localization signals interact with nuclear pore complex or, possibly, cytoplasmic components.

II. Experimental Approaches

II-A. In vivo

The early studies of nuclear import relied heavily on in vivo approaches. In vivo approaches have generally consisted of delivering a wild type or mutant nuclear protein to the cytoplasm of an appropriate host cell and subsequently determining its cellular location. Delivery methods include transfection, transformation, and microinjection. Methods for detecting a protein's cellular location include indirect immunofluorescence, histochemical stains, subcellular fractionation, electron microscopy, and functional assays.

A nuclear import substrate is frequently introduced into the desired cells by transfection. In higher eukaryotes, transient transfection assays are most commonly employed [51,59,75,117,151,161]. In studies performed with yeast, plasmid DNA encoding a nucle protein of interest is introduced by transformation a stably transformed cells are isolated by the ability plasmid borne markers to complement auxotrophic m tations [9,64,91,99,100,104,136,145]. Less frequentl nuclear import substrates have been delivered by e pression during viral infection [79,83,155].

Microinjection is an alternative route for introduci nuclear import substrates. In several cases, cloned DN has been microinjected into the nucleus to serve template to produce the desired substrate [26,76]. Th method allows one to readily assay a variety of mutan without purifying the substrates. More often, protei are microinjected. Proteins used in microinjection e periments have been isolated from the nucleus or cyt plasm of the original organism [10,29,32,33,60] or fro *E. coli* [16,93,124], have been translated in vitro [23] have been synthesized by chemically cross-linki nuclear localization signal peptides to a carrier prote (usually serum albumin) [18,56,84,85,86,152,159]. The microinjection experiments have shown that a nucle protein does not lose import competence, e.g., by pr teolytic removal of a signal, nor does it need to b synthetically modified in the eukaryotic cell to be in ported. A variety of cell types, including monkey cel [75,83,93,121], rodent cells [124, 152], human cel [18,159], Kangaroo cells [15] and *Xenopus* oocyt [16,23,26,33], have been amenable to microinjection f nuclear import studies. Advantages of microinjectio are that it does not depend on a developed transform tion system nor does it require a cloned gene.

Another common substrate which has yielded in portant information in microinjection experiments ha been protein-coated gold particles [3,34,35,40,105,121 Studies with protein-coated gold particles have show very incisively that proteins enter the nucleus throug the nuclear pore complex [40].

Following delivery of a nuclear import substrat various detection methods allow one to ascertain th cellular location of the protein being studied. Detectio methods, however, are generally constrained by havin to look directly at the nuclear and cytoplasmic compart ments of the cell; one does not have the luxury of bein able to assay nuclear localization by a molecular weigh shift in a gel since there is no localization-associate processing event (proteolytic processing or glycosyla tion) in nuclear import. A flexible and widely-use detection technique is indirect immunofluorescenc (IIF). For peptides too short to be detected by antibod ies, or for proteins with no available anti-sera, gen fusions have been commonly constructed with a numbe of different antigenic tags. These fusions serve severa purposes in addition to providing an antigenic tag; they can add to the import substrate a large cytoplasmic protein that blocks nuclear entry by simple diffusion, they can increase the stability of short peptides and

rotein fragments and they can provide an easily assayable enzymatic activity. Gene fusions have most requently been constructed with the *E. coli* enzyme l-galactosidase [16,64,76,91,99,100,118,124,136,145]. usions to pyruvate kinase [25a,76,94a,122], α-globin 6], galactokinase [94], dihydrofolate reductase and hloramphenicol acetyltransferase [94a] have also been escribed. These fusion proteins are usually detected by IF using antibodies directed against the antigenic tag rotein, although, in one case, β-galactosidase fusion roteins were also localized by a histochemical assay 24]. Methods to quantify the immunofluorescent signal in different compartments of the cell have also been cently developed [63,106]; this is particularly useful r comparing localization of the same antigen in different mutant backgrounds. Fluorescence microphotolysis as been used to monitor the actual movement of roteins into the nucleus [116].

Often, radiolabeled proteins have been microinjected to *Xenopus* oocytes and their cellular location has en traced by autoradiography of oocyte sections; ternatively, they have been detected by manual separation of nuclei and cytoplasm and subsequent quantiation of the labeled protein in each fraction by scinlation counting or gel electrophoresis and autoradiography [10,16,26,29,33,56,60]. Manual enucleation to termine the cellular distribution of a protein is common practice with *Xenopus* oocytes due to the ease with hich this can be done with these large cells. More nventional subcellular fractionation techniques have en used with yeast cells, but less reliably [64,103]. In ast cells where injection of a labeled protein is not ssible, the protein being studied is usually detected by enzymatic activity. In experiments in which proteinated gold particles are injected, the gold particle is tected by electron microscopy [3,34,35,40,105,121].

Cellular location can also be indirectly inferred from nctional assays. These include viral replication, virus aque formation [75,83,132], regulation of gene expresn [62,64,135] and ribosomal assembly [145].

-B. *In vitro*

The recent advent of in vitro systems for nuclear otein localization has made it possible to ask previsly technically impossible questions. One of the most oductive in vitro approaches has capitalized on the ique and unusual ability of *Xenopus* egg extracts to constitute nuclei [107,108]. Isolated *Xenopus* or mamalian nuclei from rat liver can be reconstituted, or ealed', by the addition of *Xenopus* egg extract. In dition, egg extracts can assemble artificial nuclei ound *Xenopus* sperm or bacteriophage λ DNA. Both e artificially assembled nuclei and the reconstituted tural nuclei exclude large non-nuclear proteins (imunoglobulin and phycoerythrin) but efficiently import fluorescently tagged nucleoplasmin or conjugates between BSA and peptides bearing the SV40 T antigen NLS [105,107]. This system has been used to show that (1) nuclear import is ATP-dependent [108], (2) binding of WGA blocks nuclear import [43], (3) nuclear import can be experimentally separated into two steps, binding followed by an ATP-dependent translocation step [105] and (4) nuclei reconstituted in the absence of WGA-binding proteins contain nuclear pores that permit the passage of dextrans (free diffusion) but no longer import proteins [42].

In vitro nuclear import systems have also employed isolated nuclei, purified from either rat liver [73,96], mouse liver [114], or *S. cerevisiae* [77,48]. The rat and mouse liver-derived in vitro nuclear import systems employ substrates (including nucleoplasmin, SV40 large T antigen, adenovirus E1b, HSV thymidine kinase, and mouse p53) biosynthetically labeled with [^{35}S]methionine or fluorescently tagged with fluorescein or phycoerythrin. The import assays consist of incubating isolated nuclei with import substrates, pelleting (in some cases through a sucrose cushion) the nuclei from this incubation mixture, and then detecting the nucleus-associated proteins by a fluorescent signal or by SDS-PAGE electrophoresis and autoradiography. The nucleus-associated proteins are judged to be inside the nucleus based on immunofluorescence [73,96], inextractability by 1% Triton X-100 [96,114] or inaccessibility to immobilized trypsin [96,114]. Furthermore, cytoplasmic variants of the SV40 large T antigen fail to be imported, demonstrating specificity for a NLS [73,96]. These in vitro systems have demonstrated that nuclear association is time-, temperature- and ATP-dependent, arguing that import is an active transport process. Interestingly, in contrast to studies using assembled or reconstituted nuclei [43,106], in some of these in vitro studies nuclear import was not enhanced by a cytoplasmic extract [96,114] or inhibited by WGA [114].

The *S. cerevisiae* in vitro import systems use substrates generated by in vitro transcription-translation of the non-yeast substrates SV40 large T antigen and nucleoplasmin [77], or the yeast proteins MCM1 and STE12 [48]. In both cases, non-nuclear proteins are excluded from the nucleus whereas nuclear proteins are imported into the nuclei as shown by co-sedimentation through a sucrose cushion and protection from externally-added protease. Import is time-, temperature- and energy (requires ATP hydrolysis) dependent. These yeast systems should further the characterization of nuclear import of the large number of yeast proteins whose import has been studied in vivo. Moreover, they should enable further characterization of nuclei from mutants that suffer an in vivo import defect (*npl1/sec63*) [131] or have an altered nuclear pore structure.

Lastly, one other in vitro approach has been the isolation and study of closed nuclear envelope ghosts

TABLE I

Nuclear localization signals (NLS)

'Position of signal' refers to the region of the protein which was shown to actually contain a NLS. NLS sequences which are derived from a large region should not be taken too literally since they are usually chosen based on homology with another NLS and therefore do not necessarily represent an actual NLS. An indicated sequence may also, depending on how it was identified, represent more than or only part of a signal.

Source	Nuclear protein	Number of amino acids	Position of signal	Deduced signal sequence	Reference
Yeast	MATα2	210	1–13	K-I-P-I-K [7]	64
			141–159	V-R-I-L-E-S-W-F-A-K-N-I [152]	63
SV40	Large T	708	126–132	P-K-K-K-R-K-V [132]	75, 76, 83
Yeast	GAL4	881	1–29	?	104, 136
Influenza virus	Nucleoprotein	498	327–445	A-A-F-E-D-L-R-V-R-S [345]	26
Yeast	Ribosomal protein L3	391	1–21	P-R-K-R [21]	100
Polyoma virus	Large T	785	183–197	V-S-R-K-R-P-R-P-A [197]	122
			near 282	P-K-K-A-R-E-D [286]	
SV40	VP1	361	1–11	A-P-T-K-R-K [6]	155
Adenovirus	E1a	289	285–289	K-R-P-R-P [289]	94
SV40	VP2 (VP3)	353	317–323	P-N-K-K-K-R-K [323]	51, 156
Frog	Nucleoplasmin	200	156–172	R-P-A-A-T-K-K-A-G-Q-A-K-K-K-K-L-D [172]	16,32,33
Rat	Glucocorticoid receptor	795	497–524	K-K-K-I-K [517]	117
			540–795	?	
Monkey	v-sis (PDGF B)	271	237–255	R-V-T-I-R-T-V-R-V-R-R-P-P-K-G-K-H-R-K [255]	87
Yeast	Histone 2B	130	28–33	G-K-K-R-S-K-A [35]	99
Chicken	v-rel	503	274–317	K-A-K-R-Q-R [303]	53
			332–503	?	
Influenza virus	NS1	237	34–38	D-R-L-R-R [38]	58
			203–237	P-K-Q-K-R-K [221]	
Frog	N1	590	531–537	V-R-K-K-R-K-T [537]	78
			548–554	A-K-K-S-K-Q-E [554]	
Human	c-myc	439	320–328	P-A-A-K-R-V-K-L-D [328]	25, 144
			364–374	R-Q-R-R-N-E-L-K-R-S-F [374]	
Human	lamin A	702	408–444	T-K-K-R-K-L-E [422]	93
HTLV-I	Rex ($p27^{n-III}$)	189	1–8	P-K-T-R-R-R-P [8]	82, 140
			11–19	S-Q-R-K-R-P-P [17]	
Adenovirus	pTP	653	362–373	R-L-P-V-R-R-R-R-R-R-V-P [373]	161
HIV-1	Tat	86	48–52	G-R-K-K-R [52]	25a, 129, 139
Frog	Lamin L_1	583	409–421	V-R-T-T-K-G-K-R-K-R-I-D-V [421]	18
Rabbit	Progesterone receptor	930	638–642	R-K-F-K-K [642]	59
Yeast	GAL80	435	1–109	?	111
			342–405	?	
HIV-1	Rev	116	38–45	R-R-N-R-R-R-R-W [45]	20, 95, 115
Yeast	SKI3	1, 432	1–296	?	120
			306–314	I-K-Y-F-K-K-F-P-K [314]	
Human	PDGF A-chain	211	196–211	P-R-E-S-G-K-K-R-K-R-K-R-L-K-P-T [211]	
Mouse	c-abl	1, 142	624–630	K-K-K-K-K [628]	147
Adenovirus	DBP	529	42–46	P-P-K-K-R [46]	101
			84–89	P-K-K-K-K-K [89]	
Yeast	TRM1	570	70–213	?	91
Chicken	c-erb-A	408	127–135	S-K-R-V-A-K-R-K-L [135]	25a
Human	c-myb	640	521–528	P-L-L-K-K-I-K-Q [528]	25a
Human	N-myc	456	337–344	P-P-Q-K-K-I-K-S [344]	25a
Human	p53	393	316–322	P-Q-P-K-K-K-P [322]	25a
Human	Hsp70	662	250–267	F-K-R-K-H-K-K-D-I-S-Q-N-K-R-A-V-R-R [267]	25a
Hepatitis B virus	Core protein	549	20–29	S-K-C-L-G-W-L-W-G [29]	112
Chicken	Ets1	441	369–388	G-K-R-K-N-K-P-K [383]	14
			388–441	?	
Yeast	Ribosomal protein L29	148	6–12	K-T-R-K-H-R-G [12]	145
			23–29	K-H-R-K-H-P-G [29]	

[123]. These vesicles rapidly import histones and nucleoplasmin (as well as the non-physiological substrate poly-lysine) while effectively excluding non-nuclear proteins. Somewhat curiously, this import was stimulated by GTP or GDP but not by any other nucleotide. These vesicles also support ATP-stimulated export of polyA containing mRNA.

II-C. Genetics

The full potential of a genetic approach in the study of nuclear import has not yet been realized. A genetic approach should provide detailed in vivo information on nuclear import and should facilitate over-production and purification of components of the import machinery for reconstitution studies. Genetic studies take two routes; isolation of the genes encoding components that play (NLS binding proteins) or may play (constituents of the NPC) a role in import, and genetic screens and selections to identify mutants defective in nuclear import.

Several features make *S. cerevisiae* the most attractive candidate for genetic analysis, not the least of which are the well developed genetics of this organism. Additionally, nuclear targeting signals have been identified in several yeast proteins (see Table I). Moreover, although few yeast proteins contain the canonical SV40 T antigen NLS, this signal can target proteins to the nucleus in yeast [9,104], arguing that some components are conserved and that findings in yeast should be generally applicable to other eukaryotes. Nuclear localization signal binding proteins have been detected in yeast [88,138], and reagents with which to clone their respective genes should soon be available. Although purification of yeast nuclei is somewhat problematic due to their small size and fragility, functional yeast nuclei can be obtained for in vitro studies [48,77]. Antibodies have been raised against isolated yeast nuclei [5,72], as well as against nuclear pore complex preparations [4]. Additionally, some antibodies against mammalian pore complex proteins cross-react with yeast nuclear envelope proteins (Ref. 6, L. Davis & G. Fink, pers. comm.). These antibody reagents provide a means to isolate genes encoding nuclear pore and envelope proteins, some of which could play a role in nuclear import; this should, in turn, allow the easy isolation of interesting mutants. The recently developed in vitro systems for import of nuclear proteins into purified yeast nuclei [48,77] should allow detailed mechanistic studies of such mutants.

The strength of yeast genetics rests not only on the ability to isolate proteins and then obtain and manipulate the desired gene, but also to design genetic screens and selections that can leapfrog many years of effort in isolating rare and difficult to manage membrane proteins of interest. Although unsuccessful, one early genetic screen to isolate mutants defective in nuclear import of ribosomal proteins proves illuminating (J. Teem & G. Fink, pers. commun.). *S. cerevisiae* is sensitive to the translation inhibitor cycloheximide; resistant mutants (*cyh2*) express a variant ribosomal protein that no longer binds cycloheximide. Importantly, resistance is recessive. Teem and Fink expressed the *CYH2* gene, encoding the cycloheximide sensitive version of the ribosomal protein, from a galactose-inducible promoter and reasoned that a cell conditionally defective (ts) in nuclear import would be unable to assemble cycloheximide sensitive ribosomes (ribosomes are assembled in the nucleus) when grown in galactose at non-permissive temperature. Mutants which failed to become sensitive to cycloheximide were obtained by this screen but all turned out to be defective in secretion (*sec* mutants); it remains enigmatic how these mutants satisfied the selection.

An alternative approach has been to fuse a NLS to an otherwise non-nuclear protein and mistarget the protein to the nucleus [131]. In this case, the SV40 large T antigen NLS or the N-terminus of nuclear GAL4 was fused to cytochrome c_1, a protein normally found in mitochondria. By a functional assay and indirect immunofluorescence, Sadler et al. [131] concluded that the hybrid proteins are now mislocalized to the nucleus. By immunofluorescence, the hybrid proteins are located at the rim or periphery of the nucleus and, by the functional assay, do not complement a cytochrome c_1 mutation and do not allow growth on glycerol. Conditional (ts) mutants that express the hybrid protein and grow on glycerol at permissive temperature were isolated. The first to be extensively characterized, *npl1*, is an allele of *sec63*, previously identified as having a role in early import into the endoplasmic reticulum (ER) [30,128].

There are three plausible explanations for the apparent role of *SEC63* in nuclear import. First, the protein may play a role in both early ER and nuclear import. *SEC63* has homology to the *E. coli* heat shock protein DnaJ, and thus its role could be to act as a chaperone to promote import. Second, because the ER is continuous with the nuclear envelope, nuclear pore components could reach their ultimate destination through the ER. A mutation affecting ER entry might thus disrupt pore assembly and consequently block nuclear import by an indirect effect. Third, from the immunofluorescence data presented, the NLS-cytochrome c_1 hybrid protein in a wild-type strain is located at the nuclear periphery. Since the lumen of the ER is continuous with the nuclear envelope, the fusion protein could actually be in the ER and not in the nucleus proper. Thus, the *sec63* mutation may simply block entry of the hybrid protein to the ER, as one would expect based on the known phenotype of *sec63*. However, this last explanation cannot easily be reconciled with the finding that a mutation in a putative NLS of

GAL4 in the GAL4-cytochrome c_1 hybrid allows this hybrid protein to reach the mitochondria.

Lastly, we note one final approach that has not yet been applied to nuclear import, genetic suppressor analysis. Here, one begins with a partially crippled protein and selects compensatory mutations that mitigate the original defect. Often these mutations identify interacting components. One could envision screening for mutations that suppress a protein bearing a defective NLS. Among such suppressor mutants, one might expect to find an altered NLS-binding protein. Similar approaches have been successfully applied to the study of protein secretion in *E. coli* [8]. However, in the case of nuclear import, this approach is complicated by the fact that, since a NLS is an integral part of a mature protein, a mutation in a NLS could easily alter more than just a targeting function. Suppressors of most NLS mutations might thus have to suppress defects in two unrelated functions and would, most likely, not be of interest.

III. Nuclear localization signals (NLS)

III-A. The signal

Proteins are targeted to the nucleus by specific signals (NLS) within the proteins' primary sequence [29]. These signals have been delineated for an expanding body of proteins (Table I), by two general methods. NLSs have been identified as sequences (1) that can by genetic or biochemical fusion render a cytoplasmic protein nuclear or (2) which when deleted or mutated, no longer promote nuclear uptake of the protein in which they reside. As has been previously shown for signals that direct proteins to other cellular compartments, there is no single, strict consensus NLS, but there are some general rules. Nuclear localization signals are: (1) typically short sequences, usually not more that 8–10 amino acids; (2) contain a high proportion of positively charged amino acids (lysine and arginine); (3) are not located at specific sites within the protein; (4) are not removed following localization; and (5) can occur more than once in a given protein.

Pioneering studies with the abundant *Xenopus* nuclear protein nucleoplasmin clearly and elegantly demonstrated that the C-terminal domain of this protein is responsible for mediating nuclear import [33]. This work showed that sequences within a mature protein direct nuclear localization and set the stage for much to follow. Studies on the yeast MATα2 protein, a transcriptional repressor, were the first to show that a nuclear protein can target a fused non-nuclear protein to the nucleus and the first to provide the actual sequence of a NLS [64]. Genetic fusions coupled with IIF and subcellular fractionation revealed that MATα2 can target the prokaryotic enzyme β-galactosidase to the nucleus and that only the N-terminal 13 amino acids of MATα2 are sufficient for this targeting. Found within this short stretch is the pentapeptide KIPIK which has been shown to be important in nuclear import [63]. A similar sequence of two positively charged amino acids flanking three hydrophobic residues, one of which is a proline, is present in several other yeast nuclear proteins and is not present in any cytoplasmic proteins currently known [61]. However, a peptide containing this sequence cannot target a crosslinked carrier protein to the nucleus in mammalian cells [18,84] nor is a homolog of this sequence in yeast histone H2B involved in nuclear import [99]. In contrast, the adenovirus pTP protein contains a MATα2 NLS homolog, RLPVR, within a sequence implicated in nuclear localization (161) and nucleoplasmin contains a similar signal, RPAATK, shown to be required for import [32].

Concurrent with the studies of MATα2, the nuclear localization signal was determined for the SV40 large T antigen. By deletions, point mutations, and gene fusions, the sequence PKKKRKV (residues 126–132) was defined as the SV40 T antigen minimal NLS [21,75,76, 83]. Subsequently, it was shown that peptides bearing this minimal sequence are sufficient to target a crosslinked carrier protein to the nucleus following microinjection [18,56,84,86]. Homologs of the SV40 NLS are found in many other nuclear proteins (Table I). Comparison of eight such homologs has generated the four residue consensus Lys-Arg/Lys-X-Arg/Lys [18]; the positive charge at the first position of this sequence (the second lysine in the SV40 T antigen minimal NLS, PKKKRKV) has been shown to be necessary for NLS function [21,84]. In general, the most conserved features seem to be several positively charged amino acids (arginine or lysine) associated with a proline. These conserved features are similar to those of the MATα2 signal, however, the T antigen signal and the MATα2 signal are historically considered to be two canonically different signals.

Although the classic studies on SV40 T antigen clearly demonstrate that PKKKRKV is a nuclear localization signal, a recent study by Rihs and Peters [124] hints at greater complexity. Rihs and Peters [124] have challenged the notion that this minimal NLS is the lone determinant of T antigen import in vivo. By measuring the kinetics of nuclear accumulation of different SV40 T antigen-β-galactosidase hybrid proteins, they found that while the canonical NLS is sufficient to mediate nuclear import over very long time periods (15 h), hybrids bearing the same sequence plus residues normally immediately N-terminal to the minimal NLS are imported at a markedly more rapid rate (within 8 to 10 min). This extra portion of the T antigen, residues 111–125 (SSDDEATADSQHSTP), is not basic and shows no homology to the minimal signal. This region may function either as a second independent NLS, or merely to

increase the efficiency of the minimal NLS. Interestingly, this region contains several amino acids known to be phosphorylated, either in the cytoplasm or in the nucleus. Although it remains to be seen if phosphorylation indeed plays a role here, these observations invite speculation about the possible role of phosphorylation in regulating nuclear import, and should foster a wide range of experiments with this and other proteins. Also, mutations in other regions of SV40 large T antigen can, curiously, inhibit nuclear import [32,151].

Although nucleoplasmin was the first protein with which it was demonstrated that a protein domain mediates nuclear import [33], identification of the actual NLS proved more exacting. By homology, four potential nuclear targeting sequences are present in the localization-promoting C-terminal tail of nucleoplasmin. Deletion analysis revealed that the sequence responsible for targeting nucleoplasmin to the nucleus is composed of two of these sequences, one (RRAATK) similar to the N-terminal NLS of MATα2 and the other (AKKKK) to the minimal NLS of SV40 large T antigen [16,32]. Whether this combination of the MATα2 and T antigen motifs constitutes one complex signal or two independent juxtaposed signals remains to be determined.

Sequences from several other proteins which have been shown or thought to confer nuclear localization are listed in Table I. In many cases the indicated sequences cannot be taken too literally since they often represent only a NLS-homologous sequence taken from within a larger domain defined by deletions as being important in mediating localization.

III-B. Dual signals

A growing number of proteins have been shown to have two independent nuclear localization signals (Table I). Among the better characterized examples of dual signals are those of polyoma large T antigen [122] and yeast ribosomal protein L29 [145]. Polyoma large T antigen contains two basic sequences (VSRKRPRPA and PKKARED) which cooperate to mediate nuclear entry. Mutation of either one individually impairs but does not eliminate the ability of polyoma virus T antigen to enter the nucleus; mutation of both results in only cytoplasmic T antigen. Similarly, ribosomal protein L29 has two basic SV40 T antigen-like signals. Either signal alone is sufficient to direct β-galactosidase to the yeast nucleus. Mutation of the C-terminal signal (KHRKHPG) has no evident effect on L29 localization; mutation of the N-terminal signal (KTRKHRG) partially impairs localization, and a combination of two mutant signals leads to an even greater loss of localization activity. These two signals are identical in five positions.

Why should a protein have two signals? Either the signals are different and have distinct, perhaps interdependent functions (like an arm and a leg), or the signals are functionally equivalent and additive (like two legs). The general assumption is that signals are equivalent and additive (see section III-C). However, in most cases of identified dual signals, the two signals are not sufficiently characterized to make firm statements about relative function.

In at least one case of a dual signal, the two signals have been proposed to be functionally distinct [63]. A second, internal signal has been identified in MATα2. This internal NLS shares no homology with either the previously identified N-terminal signal of MATα2 or other identified NLSs (Table I). It lies within the homeodomain of MATα2, suggesting that other homeodomain-containing proteins may bear homologous NLSs, perhaps evolved to coordinate nuclear import with DNA binding. The finding that MATα2 has two non-homologous signals combined with the finding that mutation of each signal individually has different effects, has led to a model in which each signal supports a different step in nuclear import [63]. According to this model, the N-terminal signal could mediate a receptor-dependent association with the nuclear pore complex and the internal signal could subsequently engage the transport machinery to foster translocation. This model provides a molecular basis for the known two steps in nuclear import, binding to the nuclear pore and subsequent ATP-dependent translocation through the pore [105, 121]. By this model, many proteins might bear two signals, or those that have only one signal might be imported more slowly.

III-C. Context and flanking sequence affect NLS function

In the studies elucidating nuclear import of the SV40 large T antigen, Kalderon et al. [75] observed that deletions to either side of the minimal NLS do not markedly affect the steady state of import. These deletions necessarily introduce new flanking sequences about the NLS. In this case, NLS function is not dramatically sensitive to the nature of abutting sequences. In later work, Roberts et al. [126] specifically addressed the role of sequence context by introducing the SV40 NLS into several different positions within the cytoplasmic enzyme pyruvate kinase. Among five constructs bearing this signal at different locations, only one insertion was unable to direct nuclear import. By extrapolating from the known three dimensional crystal structure of cat muscle pyruvate kinase to the chicken enzyme, Roberts et al. [126] found that four of the fusion proteins have their NLS on exposed surfaces whereas the sole nonfunctional position corresponds to a buried hydrophobic domain in which the NLS may be masked and thereby inactivated.

In one other case of context effects on NLS function,

Nelson and Silver [104] fused the GAL4 and V40 T antigen NLSs to both invertase and β-galactosidase. In this case, both signals drive more efficient import of invertase than of β-galactosidase. They conclude that context affects the presentation of both nuclear localization signals. However, this conclusion is tempered by the possibility that other features are likly to render some proteins less facile at nuclear impo/ than others. In this case, for example, β-galactosidase is a tetramer, and one might expect a dimer of invertase (approx. 120 kDa) to win the import race when competing against a β-galactosidase tetramer weighing in at approx. 450 kDa.

III-D. Weak signals are additive

A single NLS can clearly target a carrier protein to the nucleoplasm. Why then do not all nuclear proteins simply carry one signal? One consideration is that since NLSs are part of the mature protein, their sequence must be compatible with the overall function of the protein. To satisfy this, one might imagine that different signals have evolved to relax the constraint that all nuclear proteins bear one highly conserved signal. Secondly, it might be unnecessary for a protein to have one strong targeting signal if several weaker ones can act together (see section III-B).

Roberts et al. [126] explicitly tested the notion that reiterations of a partially defective NLS could be sufficient to drive import. The weakly active K128R variant of the SV40 T antigen NLS was fused once, twice, or thrice to pyruvate kinase. One signal yielded equal amounts of nuclear and cytoplasmic localization, two produced greater nuclear localization than cytoplasmic, and three proved sufficient to give exclusively nuclear localization. In further support of this, Lanford et al. [84] found that the inability of peptides bearing a defective SV40 T antigen signal to target carrier mouse IgG to the nucleus was mitigated when conjugates contained more peptides. Moreover, when Dworetzky et al. [35] studied import of colloidal gold particles decorated with BSA coupled to peptides bearing the wild type SV40 T antigen NLS, they observed that as the number of cross-linked peptides increased, the relative uptake of particles and the functional size of the pore channel also increased. There are also other examples in which reiteration of a NLS leads to faster more pronounced nuclear import [33,44,85]. Clearly, these findings support a model in which the effects of partially functional signals are additive. Operationally, this means that a given protein may have no single strong consensus match to any other NLS (such as the classical SV40 signal), but rather several poor matches that act additively to promote nuclear import. Thus it is impossible to assign a priori a role to any single sequence. Additionally, in these cases it should prove impossible to find by homola NLS sequence, and difficult experimentally to define which parts of the protein are functional. These may be the reasons that all known studied examples of nuclear imported proteins have at most two signals. For any protein with more than two signals, it may simply prove have proved intractable to define the regions active import.

III-E. Competition between nuclear and non-nuclear localization signals

In number of either natural or artificial cases, proteins have different signals for targeting to different organelles. What happens when a protein bears a NLS and a second signal for targeting to another cellular compartment?

Mitochondrial vs. nuclear localization signal Cytochrome c_1 is a mitochondrial protein that bears a typical N-terminal mitochondrial targeting signal. When the SV40 T antigen NLS is fused to the N-terminus of cytochrome c_1, the hybrid protein is directed instead to the nucleus, based on immunofluorescence localization and loss of cytochrome c_1 function [131]. In contrast, when the yeast nuclear protein HAP2 is fused to the C-terminus of the mitochondrial protein HEM15, this chimera is localized to mitochondria (M. Haldi and L. Guarente, pers. comm.). These findings suggest that when a protein carries both mitochondrial and nuclear localization information, the most N-terminal signal dominates. Because mitochondrial import could be cotranslational, proteins bearing the mitochondrial leader sequence at the N-terminus may initiate mitochondrial import before the NLS has been translated. One might also expect that a nascent NLS could be recognized by a signal binding protein and that nuclear import would begin prior to translation of a more C-terminal mitochondrial signal.

The yeast TRM1 protein, an enzyme that modifies both mitochondrial and cytoplasmic tRNAs, is located in both the nucleus and mitochondria. The N-terminus of TRM1 (residues 1–48 or 17–48) is sufficient to target cytochrome oxidase subunit IV or dihydrofolate reductase to mitochondria [36]. In contrast, a longer region (1–213) targets β-galactosidase to both nuclei and mitochondria [91]. Shorter fusions to β-galactosidase (1–70 or 17–70) are not nuclear localized and surprisingly, cause the loss of mitochondrial DNA, rendering it problematic to show that the protein colocalizes with DAPI-stained mitochondria. The present model is that a N-terminal mitochondrial signal and a NLS, that lies somewhere between residues 70–213, are both functional in targeting to their respective destination [91]. Thus, in at least this case, one protein can be targeted to two different locations by two independent signals.

Secretory vs. nuclear localization signal. In three examples where a protein bears both a secretory signal

sequence and a nuclear targeting signal, an influenza virus hemaglutinin-SV40 large T antigen hybrid [133], the 78 kDa glucose regulated protein [102], and platelet-derived growth factor (PDGF) [87,94a] the protein enters the ER rather than the nucleus. In all cases, the export signal sequence is N-terminal and precedes the NLS or the presumed NLS. Thus, as in the case with mitochondrial targeting, these proteins can be co-translationally inserted into the ER such that transfer is initiated prior to synthesis of the NLS. The two transport machineries would, therefore, never be directly competing for substrate.

In a more complex example, the hepatitis B virus precore protein, the protein's N-terminal signal sequence initiates ER import, but following signal peptidase cleavage translocation is aborted [112]. The processed protein is then released or leaks back into the cytoplasm. The NLS then directs nuclear import. An interesting possibility in this case is that a cytoplasmic NLS-binding protein is responsible for arresting transport into the ER.

Membrane anchor vs. nuclear localization signal. Polyoma virus middle T antigen bears a C-terminal hydrophobic domain that binds the protein to the plasma membrane. Roberts et al. [126] inserted the SV40 large T antigen NLS into polyoma virus middle T antigen. The resulting hybrid was completely cytoplasmic. When the C-terminal hydrophobic domain was deleted, this truncated hybrid entered the nucleus, arguing against the possibility that the NLS was inactive in the intact hybrid because it was not exposed on the protein surface. These findings suggest that a plasma membrane hydrophobic anchor can over-ride a NLS.

III-F. Subnuclear localization

The nucleus is arguably the largest organelle in the cell. While most nuclear proteins need carry an address that directs them only to cross the nuclear envelope, we expect some must carry a more detailed address that sends them to their ultimate intranuclear destination. Recent studies provide detailed views as to how proteins are targeted to the inner nuclear membrane and to the nucleolus.

The structural integrity of the nucleus, as well as its assembly following mitosis, depends on the lamina, a fibrillar network composed of the nuclear lamins [49]. The function of the lamins requires their intimate association with the inner membrane, where they form a supporting scaffold. Lamins A and B both have a CAAX sequence motif at their C-termini which leads to a series of processing events that endow the lamins with hydrophobic membrane anchors or targeting signals [37,70,81,149]. The cysteine of the CAAX box is farnesylated, the three C-terminal residues are proteolyzed, and the C-terminus is carboxymethylated [54,57]. These modifications are thought to direct the proteins to the inner membrane. It is not yet clear if CAAX processing occurs in the cytoplasm, although this seems likely since plasma membrane associated (RAS, γ-subunits of G-proteins) and secreted proteins (fungal mating factors) are also subject to CAAX box processing. Once bound to the inner nuclear membrane, lamin A undergoes further proteolytic processing that removes 2–3 kDa from the C-terminus, including the CAAX modifications [149], and lamin B associates with a specific lamin B receptor which could potentially recognize the farnesylated C-terminus [153]. Thus, in both cases the CAAX modifications may, in the end, not function as hydrophobic membrane anchors, but might only direct the lamins to the membrane where lamin-lamin or lamin-receptor interactions bind the lamins to the inner membrane.

Recent studies have revealed that several retroviral regulatory proteins, Rex of HTLV-1, Rev of HIV, and Tat of HIV, residue at the nucleolus [20,25a,82,139,140]. Many of the techniques originally used to study nuclear localization have been employed to address the question of how proteins are sent to the nucleolus. Point mutations and deletions were first used to implicate specific sequences in nuclear/nucleolar targeting. Subsequently, fusions were constructed to β-galactosidase, pyruvate kinase or, in one case, to another retroviral protein, $P40^x$, that is already nuclear [25a,140]. The view that has emerged is simple and elegant. Tat, Rex, and Rev all carry a basic nuclear targeting signal. For Rex (PKTRRRP), this sequence is somewhat homologous to the SV40 NLS (PKKKRKV), while those of Tat (GRKKR) and Rev (RRNRRRRW) are less similar. For each of these proteins, an additional flanking sequence appears to modify the address from nuclear to nucleolar. As few as 12 (Tat), 16 (Rev), or 20 (Rex) amino acids can serve as a nucleolar localization sequence (NuLS), i.e., can direct a passenger protein from the cytoplasm all the way to the nucleolus. Point mutations within these short sequences have further delineated the NuLS. Based on only these three proteins, a rough consensus of the additional sequence, or 'subsignal', which modifies a NLS to convert it into a NuLS is a glutamine residue flanked on both sides by several basic residues, usually arginines, i.e., **RRRQRRR**. This sequence is adjacent to or even overlapping the NLS which it modifies. Tat (GRK**KRRQRRR**) and Rex (PKTRRR**PRRSQRKR**) have one copy of the subsignal, in both cases located on the C-terminal side of the NLS. Rev (**RQARR**NRRRR**WRERQR**) appears to have two less conserved subsignals, one on each side of the NLS. Perhaps in this case, the two subsignals are weak individually and therefore two are required to properly modify the NLS. The finding that nucleolar localization signals are modified NLSs raises interesting questions as to their mechanism of action. Is the NuLS simply an

NLS in which the interaction with a binding protein is altered such that the binding protein is not disengaged until the nucleolus has been reached? Alternatively, there could be a second receptor, either in the cytoplasm or the nucleus, which recognizes the NuLS and which could functionally interact with the NLS binding protein. Since the NuLS is a modified NLS, it seems unlikely that it acts simply by binding to a nucleolar component after random diffusion into the nucleolus.

A number of other proteins have other specific subnuclear locations. In yeast, for example, the tRNA modifying enzyme TRM1 (N^2, N^2-dimethyl guanosine tRNA methyltransferase) [91] and RNA11 [17], a component of the splicesome required for mRNA splicing, are located at the nuclear periphery, within 300 nm of the nuclear envelope. tRNA ligase is found at two subnuclear locations, the extreme periphery of the nucleus at what appear to be nuclear pores and at spots more centrally located, approximately 100 nm from the envelope [19]. Clark and Abelson [19] suggest that the ligase associates with nascent tRNA's in the more central spots and then migrates, as an RNA-enzyme complex, to the pore where the entire processing machinery is located. In amphibian and insect oocytes, snRNPs are found in a newly described intranuclear structure called the sphere organelle [47]. Sphere organelles, of which there are a few dozen per nucleus, are located either extrachromosomally or in association with specific chromosomal loci called the sphere organizers, by analogy with nucleoli and the nucleolar organizing center. Gall and Callan [47] suggest that the spheres assemble snRNPs in much the same way that nucleoli build ribosomes. In somatic cells, snRNP's are found, by indirect immunofluorescence and 3D reconstruction, in 20 to 50 irregularly shaped speckles connected by a reticulum which forms a network extending throughout the nucleoplasm [46,90,142]. How any of the above proteins or particles find their specific subnuclear locations, if not by random diffusion, is unknown.

IV. Cellular components

IV-A. Nuclear pore complex (NPC)

Proteins enter [35,40] and RNA exits [34] the nucleus through the nuclear pore complex. A NPC not only supports the simultaneous [34], bidirectional flow of these two macromolecules but must do so under very heavy traffic conditions in a rapidly growing cell. The imposing structure of the NPC is, thus, not so surprising. The structure of the NPC, from amphibian oocytes, has been determined by high resolution electron microscopy combined with image processing [2,119,146]. The NPC is composed of two coaxial rings connected to a central plug by spokes. The two rings are on the surfaces of the nuclear envelope, the so-called outer ring on the cytoplasmic surface and the inner ring on the nucleoplasmic surface. As viewed down the central axis of the rings, perpendicular to the plane of the nuclear envelope, the NPC has eightfold symmetry highlighted by eight spokes which extend from the outer edge of the pore complex to a central structure. This central structure is variously referred to as the central plug, central granule or transporter. As implied by its name, the transporter is thought to contain the transport channel (see section IV-C). The NPC has an overall diameter of approx. 140 nm; the rings have an inner diameter of approx. 80 nm and the central plug has a diameter of approx. 35 nm. The density of NPC's in the nuclear envelope is usually proportional to the actual or, as in the case of the *Xenopus* oocyte, the anticipated metabolic activity of the cell [98]; the metabolically inactive lymphocyte has approx. 3 pores/μm^2, whereas the *Xenopus* oocyte has approx. 50 pores/μm^2.

The structural complexity of the NPC is often underappreciated. Mass determination of the NPC by high resolution scanning transmission electron microscopy has yielded a molecular mass of approx. 124 MDa [119], approx. 40-times the mass of a ribosome. The NPC could thus easily consist of over 100 different polypeptide species. Despite the large size and structural complexity of the NPC, only a handful of constituent proteins have been identified [22,27,28,50,141,154]. Many, if not all, of these proteins are glycoproteins containing O-linked *N*-acetylglucosamine (GlcNac) residues (for review see Ref. 68). Whether any of these proteins are actually part of the import machinery remains to be determined.

In addition to actively transporting proteins and RNA, the NPC also allows passive diffusion of small molecules, including proteins, across the nuclear envelope (113, for review see Ref. 11). The exclusion size limit for diffusion of a globular protein is approx. 60 kDa, although diffusion becomes rate-limited at approx. 20 kDa. The physiological significance of this passive diffusion is not known. The site of the passive diffusion channel in the NPC and its relationship to the active transport channel are also not known.

IV-B. Receptor (NLS-binding proteins)

The existence of receptors for nuclear targeting signals has been inferred from the specificity and saturability of nuclear protein import [61]. While the data on saturability of nuclear protein import are conflicting, the evidence for specificity is clear cut.

Nuclear import is a highly specific process. Under physiological conditions, the nucleus has a very different protein composition compared to the cytoplasm. This is accomplished despite the mixing of cytoplasmic and nuclear contents that takes place during each mitotic cycle in higher eukaryotes, when the nuclear envelope

breaks down at the beginning of prophase. Early experiments in which labeled proteins were injected into frog oocytes showed that the nucleus readily discriminates between nuclear and cytoplasmic proteins [10]. The subsequent discovery of nuclear localization signals, in which single amino acid substitutions impair function [75,83], showed that a specific sequence in a nuclear protein is the basis of this discrimination. The conclusion from these results is that specific recognition mechanisms are involved in nuclear protein localization, implying the existence of receptors for nuclear targeting signals.

Besides specificity, the other hallmark of receptor-mediated processes is saturability. Nuclei seem to have a large capacity to accumulate nuclear proteins; therefore, one would not expect to reduce the final extent of accumulation of a given protein by adding an excess of another, even if they use the same receptors. On the other hand, the rate of accumulation should be reduced by an excess of competitor protein or peptide if there is a limited number of common receptors. Analysis of the localization of SV40 T antigen signal peptide-BSA conjugates injected into *Xenopus* oocytes has provided evidence that the rate of nuclear import is saturable. These conjugates are imported into nuclei with saturable kinetics (approximate K_m of 1.8 μM and a V_{max} of $6.4 \cdot 10^9$ molecules/cell min) [56]. Furthermore, the import rate of the conjugates is reduced by millimolar concentrations of free wild type signal peptide but not by a mutant (import deficient) signal peptide. In a rat cell free system [96], a 50-fold excess of cold T antigen reduces the nuclear accumulation of labeled T antigen, although 10 mM free signal peptide, surprisingly, had no effect.

The existence of a limited number of saturable binding sites has been inferred from studies with a mutated SV40 large-T antigen that is cytoplasmic despite having a wild-type nuclear targeting signal. This protein interferes with the nuclear localization of wild-type T antigen and other nuclear proteins [132]. The interference is eliminated when the nuclear targeting signal of the cytoplasmic mutant is inactivated. This can be interpreted as evidence for a saturable number of receptor sites; the cytoplasmic mutant presumably binds an import receptor which, because the mutant is not translocated into the nucleus, remains occupied and unavailable for other proteins. This SV40 large-T antigen mutant does not interfere with the nuclear localization of all nuclear proteins tested [132], suggesting that the observed interference does not involve a binding site common to all nuclear proteins. Similar findings and conclusions are reported in a study on a mutant ICP4 protein of herpes simplex virus. A mutation in ICP4 inhibits entry of this and other, but not all, viral nuclear proteins [79]. Additional evidence for the involvement of a saturable binding site prior to translocation has come from analyzing the localization of small nuclear and non-nuclear proteins injected into the cytoplasm of cultured cells [15]. The nuclear protein histone H1 and SV40 T antigen signal-cytochrome C conjugates are retained in the cytoplasm at 4° C, whereas non-nuclear proteins of similar size diffuse into the nucleus. The cytoplasmic retention of histone H1 is overcome by injection of excess H1. This saturable cytoplasmic retention suggests that there are cytoplasmic receptors for nuclear localization signals that remain 'frozen' in place at low temperature. The excess H1 is presumably free to enter the nucleus by diffusion.

Although the above results indicate the involvement of a saturable signal recognition step, conflicting data on the saturability of nuclear import have been obtained. Others have found, in experiments involving injection into cultured cells, that millimolar concentrations of signal peptides do not reduce the rate of import of nanomolar concentration of SV40 signal-peptide-phycoerythrin conjugates [152]. The difficulty in obtaining coherent results from experiments measuring import kinetics probably stems from the technical complexity of such experiments.

The lack of a consensus sequence among the nuclear targeting signals identified so far (Table I) makes specific recognition by the translocation machinery difficult to explain. Three hypothetical mechanisms can be envisioned. First, there could be a receptor for each signal sequence. This would require a large number of different receptors on a limited number of pore complexes. Second, there could be a single receptor with broad specificity. This is hard to reconcile with the observed high specificity of nuclear protein localization, unless that which is recognized is not the primary structure of the signals but some other common determinant, e.g., a secondary structure motif. Third, there could be adaptor molecules in the cytoplasm that bind individual signals and bring them to a common receptor on the nucleus. The receptor would then recognize a determinant present only in the complex between the nuclear protein and the adaptor molecule. Of course other intermediate possibilities exist and there is little evidence to support any model at this stage. But the data collected so far tend to support the adaptor model. For example, signal-binding proteins have been described to be present in the cytoplasm, nuclear envelope and nucleoplasm (see section IV-B below), and it is known that colloidal gold particles coated with different nuclear proteins, T antigen and nucleoplasmin, are imported through the same pores, arguing against the presence of different receptors in different pores [35].

Putative receptors for nuclear localization signals have been detected (Table II) by two different techniques, direct visualization by electron microscopy of in situ binding and cell free binding studies. The probes used in the cell free binding studies have been either

TABLE II

NLS-binding proteins

Proteins are designated by their molecular weight. OGP, octylglucopyranoside; NP, nucleoplasmin.

Source	Probe	Method	Binding proteins	Extractability (dedd location)	Reference
Frog	antibody	immunoblot	p180 & p60	(nucleus)	38
Rat	antibody	immunoblot	p69 & p59	(nucleus)	158
		immunoprecip.	p65, p54, p50, p43 & p34	1.5% OGP/0.5 M (envelope)	
Rat	signal peptides	crosslinking	p66 & p56	1% OGP/0.3 M s(envelope)	1
				(cytoplasm & nucplasm)	
Rat	signal peptides	crosslinking	p138	0.15 M salt (nuc[5])	157
			p53	1 M salt (nuclei	
			p98 & p68	(cytoplasm)	
Yeast	signal peptides	ligand blot	p70 & p59	0.5 M salt but[1] 2% TX-100	138
	NP peptides	ligand blot	p140, p95, p70 & p59	(nucleus)	
Yeast	signal peptides	ligand blot	p67	(nucleus)	
				8 M urea or[6] TX-100/2 M salt	88
Rat	signal peptides	crosslinking	p74, p65, p57 & p56	(envelope)	
Human	signal peptides	crosslinking	p66	1% TX-10(envelope)	7
				sonicatio(nucleoplasm)	92

signal-bearing peptides or antibodies, although a recent report on the inhibition of binding by NEM [106] could provide another valuable tool for defining binding sites.

EM studies Evidence that nuclear proteins have discrete binding sites on the nuclear envelope has been provided by electron microscopic studies on frog oocytes injected with nucleoplasmin-coated gold particles. Early experiments showed that the injected particles are not free to diffuse immediately before or after translocation through nuclear pores, but are kept instead bound to structural elements on both sides of the pore [40]. Subsequent work extended these observations by showing that nucleoplasmin-coated colloidal gold binds to 3 nm fibril on the cytoplasmic side of the pore [121]. Binding to the octagonal array of spokes and the central plug or transporter of the pore complex (see section IV-A) has also been observed [3]. The role of these different binding sites is uncertain; they could act sequentially. The cytoplasmic filaments could bind the import substrates first, conferring vectorial movement toward the nuclear pore where they could be transferred to docking sites in the central transporter of the pore complex [3].

Despite the apparent role for cytoskeletal elements in nuclear protein import, disruption of the cytoskeleton with colcemid plus cytochalasin E does not affect the nuclear localization of an endogenous heat-shock protein or an injected adenoviral protein E1a [150]. However, the effect of this treatment on the karyoskeleton or on the pore-associated fibrils is unknown.

Cell-free binding studies. A number of nuclear pore complex proteins have been identified with lectins or antibodies raised against nuclear envelopes [27,50,66, 141]. However, we will discuss as putative import receptors only those proteins which are either recognized by antibodies that inhibit nuclear import, or which bind previously characterized nuclear targeting signals with some specificity. Proteins that meet these criteria have been found in different biological systems (see Table II). In the following discussion and in Table II, the molecular mass of the proteins identified with the aid of crosslinkers do not include the mass of the crosslinked ligand, therefore direct comparisons can be made with the proteins detected with antibodies. Furthermore, it is important to note that the physiological relevance to nuclear import of the binding proteins identified so far remains to be determined. However, the identification of any proteins at all is a promising development.

Most studies on signal-binding proteins have been performed with rat hepatocytes, and chemical crosslinking has generally been used to detect the association of peptide ligands to putative receptors. In the first such study, Adam et al. [1] describe the binding of radiolabeled peptides bearing the SV40 large-T antigen signal to two proteins of approx. 56 and 66 kDa in rat liver homogenates. The two proteins were found in the cytoplasm, the nuclear envelope and the nucleoplasm. Interestingly, the binding activity of the nuclear envelopes had to be solubilized with detergent and high salt for optimal detection; Adam et al. (1) speculate that this might be because the proteins are associated with endogenous ligands. The binding did not require ATP and was competed by unlabeled peptides with the wild-type sequence but not by mutant peptides unable to direct nuclear targeting, indicating that the binding is specific. Binding of the signal peptide to p56 was analyzed in more detail and was found to be saturable at about 80 nM peptide, regardless of the original cellular location of the p56. The different cellular locations of both p56 and p66 suggests that these proteins serve as carriers of nuclear proteins from cytoplasm into the nucleoplasm.

In a similar study with rat liver cells, Yamasaki et al.

[157] used peptides bearing the nuclear targeting signals of five different nuclear proteins as ligads [157]. It was found that each peptide has a unique pattern of affinities toward four different binding proteins. Two of the proteins (58 and 98 kDa) are cytoplamic and the other two (53 and 138 kDa) are loosely nulear, as shown by their extractability with salt or nuclease digestion. There was no unique binding protein for any individual signal, but the relative affinities of each peptide for the four proteins differed. However, in general there is little correlation between the binding affinities of signal peptides toward specific proteins as measured in this study, and the ability of these sigids to target proteins to the nucleus in vivo or in vitro. In the case of two peptides derived from the nucleoplasmin signal, one with a complete targeting signal had little, if any, affinity for the binding proteins, but it efficiently targets IgG conjugates to the nucleus in a microinjection assay using rat cells [84]. Conversely, a signal peptide with a truncated nucleoplasmin signal bound efficiently to the binding proteins, but was much less efficient at targeting conjugates. A nuclear targeting signal from the yeast protein MATα2, which is capable of targeting β-galactosidase to the yeast nucleus in vivo [64] but not conjugated proteins upon injection into mammalian cells [18,84], did not show significant affinity for four of the five binding proteins. A peptide bearing a mutant SV40 T antigen signal, which functions inefficiently in nuclear targeting in vivo and in vitro (Ref. 84, for review see Ref. 125), had almost the same affinity pattern for the binding proteins as the wild-type version of the same signal, suggesting that these binding proteins do not include those described by Adam et al. [1]. Although the reason for the discrepancy between binding affinity and localization ability is not clear, Yamasaki et al. [157] emphasize that the conformational requirements for in vitro binding may differ from those required for nuclear import in vivo. None of these proteins bind WGA (see section IV-C).

Benditt et al. [7] have described additional signal-binding proteins, also from rat hepatocytes, using purified nuclear envelopes as the starting material. The ligand employed in this study was a radiolabeled peptide containing the SV40 T antigen signal coupled to a photoactivatable crosslinker. The nuclear envelopes were extracted with 1% Triton X-100 and the extract was incubated with the ligand. Binding proteins present in the extract were covalently crosslinked to the radiolabeled ligand by UV irradiation. Four major protein species were labeled in this manner; they have apparent molecular masses of 56, 57, 65 and 74 kDa. A 100-fold excess of unlabeled ligand reduced the crosslinking with the labeled peptide by 60%, suggesting that crosslinking was a consequence of specific binding; a targeting-defective mutant peptide was not tested as a competitor. Two of the proteins identified in this study (56 and 65 kDa) could correspond to the binding proteins of 56 and 66 kDa described by Adam et al. However, it is unlikely that the 65 kDa protein, an [illegible]pe protein, corresponds to the cytosolic 68 kDa p[illegible] described by Yamasaki et al. [157] (see Table II).

Under the simple rationale that the receptor for a positively charged signal has a continuous stretch of negatively charged amino acids, Yoneda et al. [158] raised antibodies against a decapeptide containing the sequence DDDED. The resulting antiserum recognizes nuclear envelope antigens in a variety of cultured cell types, as revealed by immunofluorescence studies. This antiserum also inhibits nuclear accumulation of nucleoplasmin and signal peptide-BSA conjugates when injected into human embryonic lung cells. In Western blots of rat liver nuclear lysates the antiserum recognizes proteins of 59 and 69 kDa, reminiscent of the proteins described by others using signal peptides as ligands (see above and Table II). Attempts to fractionate the nuclear lysate and localize the antigens more precisely have thus far been unsuccessful, even though a high salt-detergent extract of nuclear envelopes can adsorb 60% of the import-inhibiting activity present in the antiserum. For unknown reasons the proteins immunoprecipitated from such an extract are not the same as those detected by Western blots, but a different set of proteins with apparent molecular masses of 34, 43, 50, 54 and 65 kDa. How these proteins are related to nuclear import is not known, but since they are recognized by an antibody that inhibits import, they could be components of the import apparatus.

Obviously the search for receptors involved in nuclear protein import is on a more secure basis when coupled with a functional assay. In this regard, frog oocytes are a valuable tool because they greatly facilitate microinjection of import substrates and reagents and the separation of nucleus and cytosol. Using this system, Featherstone et al. [38] have found that a microinjected monoclonal antibody raised against rat liver nuclear envelopes inhibits nucleoplasmin import and RNA export. The passive diffusion of non-nuclear proteins is not affected, indicating that only active import is inhibited. When used as an immunofluorescence reagent, the antibody decorates the oocyte nuclear envelope but, surprisingly, not the nuclei of the surrounding follicle cells. This suggests that the antigen may be present only in some cell types or only under certain physiological conditions. Western blot analysis of total oocyte nuclear proteins reveals antibody-binding components of 60 kDa and 180 kDa. These proteins are therefore candidates for components of the import apparatus in the oocyte nucleus. It should be noted that the antibody used in these studies does not recognize the nuclear targeting signal-binding proteins identified in rat liver cells by the same group [1].

HeLa cell nuclear proteins have also been probed

with nuclear targeting signal-bearing peptides. In this case the ligand was labeled not in the peptide moiety but in an attached photoactivatable, cleavable crosslinker. This technique allows the transfer of the label from the ligand to the receptor upon cleavage of the crosslinker. Using this approach, Li and Thomas [92] found that the SV40 T antigen signal peptide is capable of transferring the label to an acidic (pI 6) protein of 66 kDa when incubated with the supernatant of sonicated nuclei, indicating that the signal binding protein is a soluble nuclear protein. An import-deficient mutant peptide fails to transfer the labeled crosslinker. To test the correlation between binding and import function an import assay based on lysolecithin-permeabilized cells was developed. Proteins that accumulate in the nucleus are capable of transferring the crosslinker to p66, whereas non-nuclear proteins do not transfer the crosslinker. In the course of these investigations, Li and Thomas [92] observed that *Staphylococcus aureus* protein A is an avidly karyophilic protein that also associates with p66, as assayed with the label transfer technique. Peptides with the SV40 T antigen signal compete with the transfer of label from protein A to p66, even though protein A does not have an SV40-like nuclear targeting sequence. This indicates that p66 recognizes at least two signals with different primary structure. This is also the only case in which a nuclear targeting signal-binding protein apparently coincides with a major spot in a Coomassie-stained gel. The rest of the described binding proteins are minor cellular components.

Yeast nuclear proteins separated by polyacrylamide gel electrophoresis and transferred to nitrocellulose membranes contain proteins of 59 and 70 kDa that bind a fragment of the yeast nuclear protein GAL4 and a SV40 T antigen signal peptide-HSA conjugate [138]. A nucleoplasmin signal peptide-HSA conjugate binds the same two proteins and two additional proteins of 95 and 140 kDa. Conjugates with a mutated SV40 T antigen signal do not bind. All conjugates compete with each other and with a H2B conjugate for binding, indicating that the two proteins are capable of recognizing nuclear targeting signals with different sequences. This lends support to the notion that nuclear receptors are not specific for the primary structure of the signal. These signal-binding proteins can be extracted by 0.5 M NaCl but not by 2% Triton X-100, indicating that they are held in the nucleus by ionic interactions.

Lee and Melese [88] have found a signal-binding protein of 67 kDa in yeast nuclear envelopes. In this case, the ligands were HSA conjugated to either the wild-type SV40 T antigen or the H2B nuclear targeting signal. Conjugates carrying import-deficient mutant peptides did not bind. Since the protein is only extractable with 8 M urea or 2% Triton X-100/2 M KCl, it appears to be part of a karyoskeletal structure, like the nuclear pore complex or the putative yeast nuclear lamina. This protein could correspond to the 70 kDa yeast protein described by Silver et al. [138].

Cellular location of NLS-binding proteins. The cellular location of the putative receptors for nuclear targeting signals is still an open question. Immunofluorescence studies with import-inhibiting antibodies place them at the nuclear periphery [38,158]. Electron microscopy using colloidal gold particles, coated with nuclear proteins or signal peptide conjugates, detects them on the pore complex [3] and its associated fibrillar arrays [40,121]. However, some of the biochemical fractionation studies described above [1,157] and other studies [15,106] suggest that the receptor is cytoplasmic. Small numbers of receptors dispersed throughout the cytoplasm would probably not be detected by microscopic techniques, thus there is not necessarily a contradiction between the studies whose findings on the cellular location of a receptor are discrepant. A unifying model is that a receptor binds a nuclear protein in the cytoplasm and delivers it to the nuclear periphery for transport into the nucleus.

Unambiguous localization of nuclear protein binding sites will probably require a combination of different experimental approaches, since both microscopic and cellular fractionation studies are subject to artifacts. Preservation of native structures is always a concern when preparing samples for microscopic examination, and in practice cellular fractionation protocols do not always achieve clean separations. For example, it is difficult to distinguish peripheral membrane proteins from cytosolic proteins, or to define the degree of extractability that separates integral membrane and karyoskeletal proteins.

IV-C. Transporter

The nuclear protein translocation machinery, or transporter, is part of the nuclear pore complex (see section IV-A), transiently associates with its protein substrates, uses ATP and is inhibited by WGA. Otherwise, the structure and function of the transporter remain largely unknown.

WGA has been the most useful tool so far in analyzing the transporter, although it may not be generally applicable since it does not inhibit nuclear import in some in vitro systems [48,77, 114]. WGA binds with highest affinity to di- and oligo-*N*-acetylglucosamine residues. Unlike other forms of glycosylation, O-glycosylation of proteins with *N*-acetylglucosamine is restricted to the cytosolic and nuclear compartments of the cell (for review see Ref. 68). There are several nuclear proteins that are O-glycosylated and bind WGA [27,28,66,69,159], but their role in protein import remains to be demonstrated. Evidence in favor of such a role is that a monoclonal anti-pore antibody, one of many which include *N*-acetylglucosamine as part of

their epitopes, inhibits nuclear import in *Xenopus* oocytes [38]. Also, WGA [24,43,134,152,159] and a lectin from *Datura stamonium* with the same specificity [159] inhibit nuclear import, while other lectins with different sugar specificities do not [43,159]. Inhibition by WGA is concentration dependent [24,159] and is reversed upon addition of excess *N*-acetylglucosamine [43,159]. Binding of this lectin to the nucleus does not affect the functional pore radius for free diffusion [24,74,159] or the mobility of nucleoplasmin in the cytoplasm [24]. WGA and nuclear proteins do not compete for binding to the nucleus [105], even though nucleoplasmin- and WGA-coated gold particles appear to bind to the same pore structures [3]. Thus, it is generally accepted that this lectin specifically blocks translocation without affecting protein binding or passive diffusion through the pore.

That WGA-binding NPC proteins play a role in nuclear protein import has been elegantly demonstrated in a recent report [42]. Import-competent nuclear envelopes assemble de novo from *Xenopus* egg extracts. When the extracts are depleted of WGA-binding proteins (major species of 60, 97 and 200 kDa) prior to nuclear assembly, the newly formed nuclear pores have normal morphology and the proper exclusion properties as measured by dextran diffusion, but are deficient for binding and import of signal peptide-HSA conjugates. Addition of the missing proteins, even from heterologous sources, restores binding and import. The finding in this report [42] that nuclei without WGA-binding proteins are not even able to bind an import substrate is surprising in light of the previous finding by the same group that WGA does not affect binding [105] but only subsequent translocation. To account for this apparent discrepancy, the authors offer the hypothesis that glycosylated receptors in intact nuclei are not accessible to WGA, or that glycosylated proteins are essential for the assembly of the receptors into the pore complex. In any case, this experimental system should be a powerful tool to identify components of the translocation apparatus.

Nuclear localization of proteins naturally present in both the nucleus and cytoplasm has recently been reported to be less affected by WGA than that of exclusively nuclear proteins [24]. If confirmed, this observation could uncover a second pathway for nuclear protein import.

Nuclear protein import is perhaps the only case of protein transport in which the physical location of the translocation machinery is known. EM studies have shown that import substrates are translocated through the nuclear pore complex [35,40,105,121]. Furthermore, high resolution analysis of EM images has resolved the binding of colloidal gold coated with nucleoplasmin and WGA to discrete elements of the pore complex [3]. The element of the NPC that appears to mediate the actual translocation across the nuclear envelope is the central plug or now-called transporter. This is a doughnut shaped structure with an overall diameter of approx. 35 nm, surrounded by a ring of electon-transparent material (possibly the passive diffusion channel). Nucleoplasmin-coated particles on their way into the nucleus appear to first bind around the center of the transporter in an octagonally symmetrical pattern, and then bind to the exact center of the transporter where there appears to be a gated channel [3]. Current models on how translocation takes place envision the central transporter of the pore complex as a diaphragm. The opening of this diaphragm presumably depends on a signal recognition even and ATP hydrolysis. The functional diameter of the channel depends, at least partly, on the strength of the interaction between the signal sequence and the receptor site, since it is affected by the type and number of signal sequences present on the substrate [35,85]. Unlike protein import into other organelles, translocation across the nuclear envelope does not require structural flexibility in the import substrates, since the pore can accommodate rigid gold particles of up to 26 nm in diameter [35] (the estimated functional diameter of the free diffusion channel is approx. 9 nm [113]. WGA and import-inhibiting antibodies which recognize O-glycosylated NPC proteins could inhibit translocation by crosslinking components of the central transporter and thereby restricting the opening of the channel [3].

V. Translocation mechanism

Nuclear import is clearly an active process, requiring both physiological temperature and ATP. At low temperature or in the absence of ATP, import substrates bind to the nuclear periphery but are not translocated across the nuclear envelope [105,121]. The requirement for physiological temperature has been demonstrated in vivo [13,121,152] and in vitro [48,73,77,96,107,114,134]. The ATP requirement has also been investigated in a number of systems. Nuclear protein import is abolished following depletion of ATP with apyrase in cell-free systems [48,77,96,105,107,108,114] or with deoxyglucose and sodium azide in cultured cells [121].

ATP is widely thought to serve as an energy source for import. However, a study analyzing the nucleotide requirements for the import of a nucleoplasmin-phycoerythrin conjugate into isolated rat liver nuclei has reached a different conclusion [73]. This study found that each of four nucleotide triphosphates (ATP, GTP, UTP and CTP) and α,β-methylene-ATP are almost equally effective in fulfilling the nucleotide requirement; thus it was concluded that the role of the nucleotide is to promote an energy-independent binding step (not the actual translocation) of the protein conjugate. The strongest argument that ATP is not an energy source was the import-promoting capability of α,β-

methylene-ATP, which was used as a non-hydrolyzable ATP analog. However, this conclusion is uncertain since this compound has a potentially hydrolyzable γ-phosphate [160]. Experiments using β,γ-imido-ATP, which lacks such a phosphate, have shown that at least in isolated yeast nuclei ATP hydrolysis is required for the import process [48].

There is only one report challenging the need for ATP. Heparin-treated envelopes from rat liver nuclei accumulate histones, nucleoplasmin and polylysine, but not immunoglobulins, myoglobin or cytochrome C. Accumulation was reported to be stimulated by GTP and GDP but not by any other nucleotide [123]. However, the significance of these results is unclear, since the histones, nucleoplasmin and polylysine accumulated in the absence of GTP and the stimulation by the nucleotide was only 30%.

VI. Regulation of nuclear protein import

Under physiological conditions, nuclear protein localization must be subject to complicated regulatory mechanisms since the presence of certain proteins in the nucleus is required only at very specific moments in the cell cycle or only in response to short-lived stimuli. A few examples of such regulation have been documented, and possible regulatory mechanisms have begun to emerge.

One way to regulate import is through phosphorylation, either of the imported protein itself or of a protein with which it interacts. A peptide sequence flanking the nuclear targeting signal of SV40 large-T antigen contains phosphorylation sites; removal of this sequence significantly reduces the import rate of T antigen [124], suggesting that efficient import of T antigen requires that it be phosphorylated (see section III-A). Nuclear localization of the yeast protein SWI5 is also regulated by phosphorylation; however, in this case, phosphorylation of SWI5 prevents import (T. Moll and K. Nasmyth, pers. commun.). Localization of the mammalian transcriptional activator NF-κB is regulated by phosphorylation of an interacting protein. NF-κB is initially located in the cytoplasm in an inactive form complexed with the inhibitory protein IκB [89]. As shown in a cell-free system, phosphorylation of IκB by cAMP-dependent protein kinase or protein kinase C releases NF-κB to move into the nucleus [52,134]. The dissociation of IκB presumably unmasks a localization signal in NF-κB.

Interaction with other proteins could be a general mechanism to keep otherwise nuclear proteins in the cytoplasm until their function is required [71]. The regulation of nuclear protein localization by interaction with cytoplasmic elements has been demonstrated for the bovine cAMP-dependent protein kinase II and has been inferred for the c-rel protein (the product of the cellular counterpart of the *v-rel* oncogene), the *Drosophila* morphogen *dorsal* and steroid receptors. cAMP-PK is a tetramer of two regulatory and two catalytic subunits normally associated with the Golgi complex. Following elevation of intracellular cAMP levels, the complex dissociates and the catalytic subunits are imported into the nucleus, while the regulatory subunits remain associated with the Golgi. When the cAMP concentration returns to basal levels the catalytic subunits re-associate with their regulatory counterparts, indicating that nuclear import of this protein is a reversible process [110]. The product of the *c-rel* gene is a cytoplasmic protein whereas the product of the *v-rel* gene is nuclear. Substitution of the carboxy terminus of v-rel protein with that of c-rel inhibits nuclear localization of the erstwhile v-rel, suggesting that the carboxy terminus of c-rel acts as a cytoplasmic anchor that prevents otherwise constitutive nuclear localization [65]. Similar, the *Drosophila* morphogen *dorsal*, which is homologous to the c-rel protein, has carboxy-terminal sequences that keep it in the cytoplasm, presumably by interactions with the products of the genes *toll* or *cact*. Release of the interaction or deletion of C-terminal sequences permits nuclear localization of the dorsal protein, thereby allowing it to function in determining cell fate along the dorsal-ventral axis [127,130,143]. Steroid receptors are inactive, cytoplasmic and complexed with HSP90 in the absence of ligand [80]. Upon receptor activation by the appropriate steroid ligand, HSP90 dissociates and the receptor moves into the nucleus. Although a role for HSP90 in keeping the receptor in the cytoplasm is not clear, it could either mask a nuclear localization signal or interfere with a possibly localization-promoting phosphorylation of the receptor. The receptor is phosphorylated coincident with HSP90 dissociation.

Although in some cases nuclear proteins are anchored in the cytoplasm by interacting proteins, in other cases interaction with a second protein (other than a component of the localization machinery) does not interfere or is actually necessary for nuclear localization. An import-deficient subunit of the homopentameric *Xenopus* nucleoplasmin can be carried into the nucleus by an import-proficient subunit [33]. Yeast histone H2B with a deleted nuclear targeting signal still becomes nuclear if its domain for interaction with histone H2A is intact, indicating that a protein can carry a different protein, as a heterodimer, into the nucleus [99]. Localization of the adenovirus 140 kDa DNA polymerase is facilitated by interaction with another viral protein, pTP, again suggesting that a protein can use another protein to 'piggyback' into the nucleus [161]. The CDC2 protein kinase and the CDC13 cyclin form a complex and are co-localized to the nucleus. In the absence of CDC13, CDC2 is not active as a kinase and is no longer nuclear, suggesting that the cyclin regulates both the enzymatic

tivity and the cellular location of the kinase [12]. inally, in some instances, it is RNA with which proins need to associate before gaining entry to the icleus. *Xenopus* U2 snRNA and its binding proteins e excluded from the nucleus if present separately, but come nuclear after assembly into ribonucleoprotein rticles [39,97].

I. Conclusion / projections

Which areas of nuclear protein import could, or ould, receive the most attention by investigators in upcoming years? Clearly, the structure, function and sembly of the transporter as well as the entire NPC is central importance. Given the importance and comxity of the NPC, our understanding of this structure extremely rudimentary. The mechanism by which clear protein localization is regulated and possibly ated to signal transduction is also intriguing. There hints in the literature [24,79,132] that nuclear import olves more than one pathway; direct evidence for a ond pathway could emerge. The issue of subnuclear alization is also in its infancy. In the near future, we uld see which, if not all, of the many NLS-binding teins identified so far are actual receptors. This uld, in turn, allow more sophisticated experiments igned to probe the function of the import machinery. should also see greater application of genetics to the blem of nuclear import.

knowledgements

We thank Kevin Baker, Maria Cárdenas and Howard zman for comments on the manuscript. J.G.-B. and I. were recipients of EMBO Fellowships. The laboray, in which some of the experiments reviewed here re performed, was supported by a grant from the iss National Science Foundation to M.N.H.

ferences

Adam, S.A., Lobl, T.J., Mitchell, M.A. and Gerace, L. (1989) Nature 337, 276–279.
Akey, C.W. (1989) J. Cell. Biol. 109, 955–970.
Akey, C.W. and Goldfarb, D.S. (1989) J. Cell Biol. 109, 971–998.
Allen, J.L. and Douglas, M.G. (1989) J. Ultrastruc. and Mol. Struc. Res. 102, 95–108.
Aris, J.P. and Blobel, G. (1988) J. Cell Biol. 107, 17–31.
Aris, J.P. and Blobel, G. (1989) J. Cell Biol. 108, 2059–2067.
Benditt, J.O., Meyer, C., Fasold, H., Barnard, F.C and Riedel, N. (1989) Proc. Natl. Acad. Sci. USA 86, 9327–9331.
Benson, S.A., Hall, M.N. and Silhavy, T.J. (1985) Annu. Rev. Biochem. 53, 101–131.
Benton, B.M., Eng, W.-K., Dunn, J.J., Studier, F.W., Sternglanz, R. and Fisher, P.A. (1990) Mol. Cell. Biol. 10, 353–360.
Bonner, W.M. (1975) J. Cell Biol. 64, 431–437.
Bonner, W.M. (1978) in The Cell Nucleus (Busch, H., ed.), pp. 97–148, Academic Press, New York.
12 Booher, R.N., Alfa, C.E., Hyams, J.S. and Beach, D. (1989) Cell 58, 485–497.
13 Borer, R.A., Lehner, C.F., Eppenberger, H.M. and Nigg, E.A. (1989) Cell 56, 379–390.
14 Boulukos, K.E., Pognonec, P., Rabault, B., Begue, A. and Ghysdael, J. (1989) Mol. Cell. Biol. 9, 5718–5721.
15 Breeuwer, M. and Goldfarb, D.S. (1990) Cell 60, 999–1008.
16 Bürglin, T.R. and De Robertis, E.M. (1987) EMBO J 6, 2617–2625.
17 Chang, T.-H., Clark, M.W., Lustig, A.J., Cusick, M.E. and Abelson, J. (1988) Mol. Cell. Biol. 8, 2379–2393.
18 Chelsky, D., Ralph, R. and Jonak, G. (1989) Mol. Cell. Biol. 9, 2487–2492.
19 Clark, M.W. and Abelson, J. (1987) J. Cell Biol. 105, 1525–1526.
20 Cochrane, A.W., Perkins, A. and Rosen, C.A. (1990) J. Virol. 64, 881–885.
21 Colledge, W.H., Richardson, W.D., Edge, M.D. and Smith, A.E. (1986) Mol. Cell Biol. 6, 4136–4139.
22 D'Onofrio, M., Starr, C.M., Park, M.K., Holt, G.D., Haltiwanger, R.S., Hart, G.W. and Hanover, J.A. (1988) Proc. Natl. Acad. Sci. USA 85, 9595–9599.
23 Dabauvalle, M.-C. and Franke, W.W. (1982) Proc. Natl. Acad. Sci. USA 79, 5302–5306.
24 Dabauvalle, M.-C., Shulz, B., Scheer, U. and Peters, R. (1988) Exp. Cell Res. 174, 291–296.
25 Dang, C.V. and Lee, W.M.F. (1988) Mol. Cell. Biol. 8, 4048–4054.
25a Dang, C.V. and Lee, W.M.F. (1989) J. Biol. Chem. 254, 18019–18023.
26 Davey, J., Dimmock, N.J. and Colman, A. (1985) Cell 40, 667–675.
27 Davis, L.I. and Blobel, G. (1986) Cell 45, 699–709.
28 Davis, L.I. and Blobel, G. (1987) Proc. Natl. Acad. Sci. USA 84, 7552–7556.
29 De Robertis, E.M., Longthorne, R.F. and Gurdon, J.B. (1978) Nature 272, 254–256.
30 Deshaies, R.J. and Schekman, R. (1987) J. Cell Biol. 105, 633–646.
31 Dingwall, C. and Laskey, R.A. (1986) Annu. Rev. Cell Biol. 2, 367–390.
32 Dingwall, C., Robbins, J., Dilworth, S.M., Roberts, B. and Richardson, W.D. (1988) J. Cell Biol. 107, 841–849.
33 Dingwall, C., Sharnick, S.V. and Laskey, R.A. (1982) Cell 30, 449–458.
34 Dworetzky, S.I. and Feldherr, C.M. (1988) J. Cell Biol. 106, 575–584.
35 Dworetzky, S.I., Lanford, R.E. and Feldherr, C.M. (1988) J. Cell Biol. 107, 1279–1287.
36 Ellis, S.R., Hopper, A.K. and Martin, N.C. (1989) Mol. Cell. Biol. 9, 1611–1620.
37 Farnsworth, C.C., Wolda, S.L., Gelb, M.H. and Glomset, J.A. (1989) J. Biol. Chem. 264, 20422–20429.
38 Featherstone, C., Darby, M.K. and Gerace, L. (1988) J. Cell. Biol. 107, 1289–1297.
39 Feeney, R.J. and Zieve, G.W. (1990) J. Cell Biol. 110, 871–881.
40 Feldherr, C.M., Kallenbach, E. and Shultz, N. (1984) J. Cell Biol. 99, 2216–2222.
41 Feldherr, C.M., Paine, P.L. and Hodges, P. (1988) Exp. Cell Res. 179, 527–534.
42 Finlay, D.R. and Forbes, D.J. (1990) Cell 60, 17–29.
43 Finlay, D.R., Newmeyer, D.D., Price, T.M., and Forbes, D.J. (1987) J. Cell Biol. 104, 189–200.
44 Fischer-Fantuzzi, L. and Vesco, C. (1988) Mol. Cell. Biol. 8, 5495–5503.
45 Franke, W.W., Scheer, U., Krohne, G. and Jarasch, E.-D. (1981) J. Cell Biol. 91, 39s–50s.
46 Fu, X-D. and Maniatis, T. (1990) Nature 343, 437–441.
47 Gall, J.G. and Callan, H.G. (1989) Proc. Natl. Acad. Sci. USA 86, 6635–6639.

48 Garcia-Bustos, J.F., Wagner, P. and Hall, M.N. (1991) Exp. Cell Res. 192, 213–219.
49 Gerace, L. and Burke, B. (1988) Annu. Rev. Cell Biol. 4, 335–374.
50 Gerace, L., Ottaviano, Y. and Kondor-Koch, C. (1982) J. Cell Biol. 95, 826–837.
51 Gharakhanian, E., Takahashi, J. and Kasamatsu, H. (1987) Virology 157, 440–448.
52 Ghosh, S. and Baltimore, D. (1990) Nature 344, 678–682.
53 Gilmore, T.D. and Temin, H.M. (1988) J. Virol. 62, 703–714.
54 Glomset, J.A., Gelb, M.H. and Farsworth, C.C. (1990) TIBS 15, 139–142.
55 Goldfarb, D.S. (1989) Curr. Op. Cell Biol. 1, 441–446.
56 Goldfarb, D.S., Gariepty, J., Schoolnik, G. and Kornberg, R.D. (1986) Nature 322, 641–644.
57 Goldstein, J.L. and Brown, M.S. (1990) Nature 343, 425–430.
58 Greenspan, D., Palese, P. and Krystal, M. (1988) J. Virol. 62, 3020–3026.
59 Guichon-Mantel, A., Loosfelt, H., Lescop, P., Sar, S., Atger, M., Perrot-Applanat, M. and Milgrom, E. (1989) Cell 57, 1147–1154.
60 Gurdon, J.B. (1970) Proc. R. Soc. B. 176, 303–314.
61 Hall, M.N. and Fried, S.R. (1988) in Molecular Biology of Intracellular Protein Sorting and Organelle Assembly (Bradshaw, R.A., McAlister-Henn, L. and Douglas, M.G., ed.) pp. 187–192, Alan R. Liss, New York.
62 Hall, M.N. and Johnson, A.D. (1987) Science 237, 1007–1012.
63 Hall, M.N., Craik, C. and Hiraoka, Y. (1990) Proc. Natl. Acad. Sci. USA 87, 6954–6958.
64 Hall, M.N., Hereford, L. and Herskowitz, I. (1984) Cell 36, 1057–1065.
65 Hannick, M. and Temin, H.M. (1989) Mol. Cell. Biol. 59, 1165–1177.
66 Hanover, J.A., Cohen, C.K., Willingham, M.C. and Park, M.K. (1987) J. Biol. Chem. 262, 9887–9894.
67 Harris, J.R. (1978) Biochim. Biophys. Acta 515, 55–104.
68 Hart, G.W., Haltiwanger, R.S., Holt, G.D. and Kelly, W.G. (1989) Annu. Rev. Biochem. 58, 841–874.
69 Holt, G.D., Snow, C.M., Senior, A., Haltiwanger, R.S., Gerace, L. and Hart, G.W. (1987) J. Cell Biol. 104, 1157–1164.
70 Holtz, D., Tanaka, R.A., Hartwig, J. and McKeon, F. (1989) Cell 59, 969–977.
71 Hunt, T. (1989) Cell 59, 949–951.
72 Hurt, E.C., McDowall, A. and Schimmang, T. (1988) Eur. J. Cell Biol. 46, 554–563.
73 Imamoto-Sonobe, N., Yoneda, Y., Iwamoto, R., Sugawa, H. and Uchida, T. (1988) Proc. Natl. Acad. Sci. USA 85, 3426–3430.
74 Jiang, L.-W. and Schindler, M. (1986) J. Cell Biol. 102, 853–858.
75 Kalderon, D., Richardson, W.D., Markham, A.F. and Smith, A.E. (1984) Nature 311, 33–38.
76 Kalderon, D., Roberts, B.L., Richardson, W.D. and Smith, A.E. (1984) Cell 39, 499–509.
77 Kalinich, J.F. and Douglas, M.G. (1989) J. Biol. Chem. 264, 17979–17989.
78 Kleinschmidt, J.A. and Seiter, A. (1988) EMBO J 7, 1605–1614.
79 Knipe, D.M. and Smith, J.L. (1986) Mol. Cell. Biol. 6, 2371–2381.
80 Kost, S.L., Smith, D.F., Sullivan, W.P., Welch, W.J. and Toft, D.O. (1989) Mol. Cell. Biol. 9, 3829–3838.
81 Krohne, G., Waizenegger, I. and Höger, T.H. (1989) J. Cell Biol. 109, 2003–2011.
82 Kubota, S., Siomi, H., Satoh, T., Endo, S., Maki, M. and Hatanaka, M. (1989) Biochem. Biophys. Res. Commun. 162, 963–970.
83 Lanford, R.E. and Butel, J.S. (1984) Cell 37, 801–813.
84 Lanford, R.E., Feldherr, C.M., White, R.G., Dunham, R.G. and Kanda, P. (1990) Exp. Cell Res. 186, 32–38.
85 Lanford, R.E., Kanda, K. and Kennedy, R.C. (1986) Cell 46, 575–582.
86 Lanford, R.E., White, R.G., Dunham, R.G. and Kanda, P. (1988) Mol. Cell. Biol. 8, 2722–2729.
87 Lee, B.A., Maher, D.W., Hannink, M. and Donoghue, D.J. (1987) Mol. Cell. Biol. 7, 3527–3537.
88 Lee, W.-C. and Melese, T. (1989) Proc. Natl. Acad. Sci. USA 86, 8808–8812.
89 Lenardo, M.J. and Baltimore, D. (1989) Cell 58, 227–229.
90 Lerner, E.A., Lerner, M.R., Janeway, C.A. and Steitz, J.A. (1981) Proc. Natl. Acad. Sci. USA 78, 2737–2741.
91 Li, J.-M., Hopper, A.K. and Martin, N.C. (1989) J. Cell Biol. 109, 1411–1419.
92 Li, R. and Thomas, J.O. (1989) J. Cell Biol. 109, 2623–2632.
93 Loewinger, L. and McKeon, F. (1988) EMBO J 7, 2301–2309.
94 Lyons, R.H., Ferguson, B.Q. and Rosenberg, M. (1987) Mol. Cell. Biol. 7, 2451–2456.
94a Maher, D.W., Lee, B.A. and Donoghue, D.J. (1989) Mol. Cell. Biol. 9, 2251–2253.
95 Malim, M.H., Bohnlein, S., Hauber, J. and Cullen, B.R. (1989) Cell 58, 205–214.
96 Markland, W., Smith, A.E. and Roberts, B.L. (1987) Mol. Cell Biol. 7, 4255–4265.
97 Mattaj, I.W. and De Robertis, E.M. (1985) Cell 40, 111–118.
98 Maul, G.G. (1977) in International Review of Cytology, (Bourne, G.H., Danielli, J.F. and Jeon, K.W., eds.), pp. 75–186, Academic Press, New York.
99 Moreland, R.B., Langevin, G.L., Singer, R.H., Garcea, R.L. and Hereford, L.M. (1987) Mol. Cell. Biol. 7, 4048–4057.
100 Moreland, R.B., Nam, H.G., Hereford, L.M. and Fried, H.M. (1985) Proc. Natl. Acad. Sci. USA 82, 6561–6565.
101 Morin, N., Delsert, C. and Klessig, D.F. (1989) Mol. Cell. Biol. 9, 4372–4380.
102 Munro, S. and Pelham, H.R.B. (1986) Cell 46, 291–300.
103 Nagy, M., Laporte, J., Penverne, B. and Hervé, G. (1982) J. Cell Biol. 92, 790–794.
104 Nelson, M. and Silver, P. (1989) Mol. Cell. Biol. 9, 384–389.
105 Newmeyer, D.D. and Forbes, D.J. (1988) Cell 52, 641–653.
106 Newmeyer, D.D. and Forbes, D.J. (1990) J. Cell Biol. 110, 547–557.
107 Newmeyer, D.D., Finlay, D.R. and Forbes, D.J. (1986) J. Cell Biol. 103, 2091–2102.
108 Newmeyer, D.D., Lucocq, J.M., Burglin, T.R. and De Robertis, E.M. (1986) EMBO J 5, 501–510.
109 Newport, J.W. and Forbes, D.J. (1987) Annu. Rev. Biochem. 56, 535–565.
110 Nigg, E.A., Hilz, H., Eppenberger, H.M. and Dutley, F. (1985) EMBO J. 4, 2801–2806.
111 Nogi, Y. and Fukusawa, T. (1989) Mol. Cell. Biol. 9, 3009–3017.
112 Ou, J.-H., Yeh, C.-T. and Yen, T.S.B. (1989) J. Virol. 63, 5238–5243.
113 Paine, P.L., Moore, L.C. and Horowitz, S.B. (1975) Nature 254, 109–114.
114 Parnaik, V.K. and Kennady, P.K. (1990) Mol. Cell Biol. 10, 1287–1292.
115 Perkins, A., Cochrane, A.W., Ruben, S.M. and Rosen, C.A. (1989) J. Acq. Imm. Def. Synd. 2, 256–263.
116 Peters, R. (1986) Biochim. Biophys. Acta 864, 305–359.
117 Picard, D. and Yamamoto, K. (1987) EMBO J 6, 3333–3340.
118 Pinkham, J.L., Oleson, J.T. and Guarente, L.P. (1987) Mol. Cell. Biol. 7, 578–585.
119 Reichelt, R., Holzenburg, A., Buhle, E.L., Jarnik, M., Engel, A. and Aebi, U. (1990) J. Cell Biol. 110, 883–894.
120 Rhee, S.-K., Icho, T. and Wickner, R.B. (1989) Yeast 5, 149–158.
121 Richardson, W.D., Mills, A.D., Dilworth, S.M., Laskey, R.A. and Dingwall, C. (1988) Cell 52, 655–664.
122 Richardson, W.D., Roberts, B.L. and Smith, A.E. (1986) Cell 44, 77–85.

3 Riedel, N., Bachmann, M., Prochnow, D., Richter, H.-P. and Fasold, H. (1987) Proc. Natl. Acad. Sci. USA 84, 3540–3544.
4 Rihs, H.P. and Peters, R. (1989) EMBO J 8, 1479–1484.
5 Roberts, B. (1989) Biochim. Biophys. Acta 1008, 263–280.
6 Roberts, B.L., Richardson, W.D. and Smith, A.D. (1987) Cell 50, 465–475.
7 Roth, S., Stein, D. and Nüsslein-Volhard, C. (1989) Cell 59, 1189–1202.
8 Rothblatt, J.A., Deshaies, R.J., Sanders, S.L., Daum, G. and Schekman, R. (1989) J. Cell Biol. 109, 2641–2652.
9 Rubens, S., Perkins, A., Purcell, R., Joung, K., Sia, R., Burghoff, R., Haseltine, W.A. and Rosen, C.A. (1989) J. Virol. 63, 1–8.
0 Rushlow, C.A., Han, K., Manley, J.L. and Levine, M. (1989) Cell 59, 1165–1177.
1 Sadler, I., Chiang, A., Kurihara, T., Rothblatt, J., Way, J. and Silver, P. (1989) J. Cell Biol. 109, 2665–2675.
2 Schneider, J., Schindewolf, C., van Zee, K. and Fanning, E. (1988) Cell 54, 117–125.
3 Sharma, S., Rodgers, L., Brandsma, J., Gething, M.-J. and Sambrook, J. (1985) EMBO J. 4, 1479–1489.
4 Shirakawa, F. and Mizel, S.B. (1989) Mol. Cell. Biol. 9, 2424–2430.
5 Silver, P.A., Chiang, A. and Sadler, I. (1988) Genes and Development 2, 707–717.
6 Silver, P.A., Keegan, L. and Ptashne, M. (1984) Proc. Natl. Acad. Sci. USA 81, 5951–5955.
7 Silver, P. and Hall, M.N. (1988) in Protein Transport and Organelle Biogenesis (Das, C. and Robbins, P.W., ed.), pp. 749–769, Academic Press, San Diego.
8 Silver, P., Sadler, I. and Osborne, M.A. (1989) J. Cell Biol. 109, 983–989.
9 Siomi, H., Shida, H., Maki, M. and Hatanaka, M. (1990) J. Virol. 64, 1803–1807.
0 Siomi, H., Shida, H., Nam, S.H., Nosaka, T., Maki, M. and Hatanaka, M. (1988) Cell 55, 197–209.
141 Snow, C.M., Senior, A. and Gerace, L. (1987) J. Cell Biol. 104, 1143–1157.
142 Spector, D.L. (1990) Proc. Natl. Acad. Sci. USA 87, 147–152.
143 Steward, R. (1989) Cell 59, 1179–1188.
144 Ueno, K., Katoh, K. and Kondoh, H. (1988) Mol. Cell. Biol. 8, 4529–4532.
145 Underwood, M.R. and Fried, H. (1990) EMBO J 9, 91–100.
146 Unwin, P.N.T. and Milligan, R.A. (1982) J. Cell Biol. 93, 63–75.
147 Van Etten, R.A., Jackson, V. and Baltimore, D. (1989) Cell 58, 669–678.
148 Verner, K. and Schatz, G. (1988) Science 241, 1307–1313.
149 Vorburger, K., Kitten, G.T. and Nigg, E.A. (1989) EMBO J. 8, 4007–4013.
150 Welch, W.J. and Feramisco, J.R. (1985) Mol. Cell. Biol. 5, 1571–1581.
151 Welsh, J.D., Swimmer, C., Cocke, T. and Shenk, T. (1986) Mol. Cell. Biol. 6, 2207–2212.
152 Wolff, B., Willingham, M.C. and Hanover, J.A. (1988) Exp. Cell Res. 178, 318–334.
153 Worman, H.J., Yuan, J., Blobel, G. and Georgatos, S.D. (1988) Proc. Natl. Acad. Sci. USA 85, 8531–8534.
154 Wozniak, R.W., Bartnik, E. and Blobel, G. (1989) J. Cell Biol. 108, 2083–2092.
155 Wychowski, C., Benichou, D. and Girard, M. (1986) EMBO J 5, 2569–2576.
156 Wychowski, C., Benichou, D. and Girard, M. (1987) J. Virol. 61, 3862–3869.
157 Yamasaki, L., Kanda, P. and Lanford, R.E. (1989) Mol. Cell. Biol. 9, 3028–3036.
158 Yoneda, Y., Imamoto-Sonobe, N., Matsuoka, Y., Iwamoto, R., Kiho, Y. and Uchida, T. (1988) Science 242, 275–278.
159 Yoneda, Y., Imamoto-Sonobe, N., Yamaizumi, M. and Uchida, T. (1987) Exp. Cell Res. 173, 586–595.
160 Yount, R.G. (1975) Adv. Enzymol. 43, 1–56.
161 Zhao, L.-J. and Padmanabhan, R. (1988) Cell 55, 1005–1015.

J Mol Evol (2003) 56:77–88
DOI: 10.1007/s00239-002-2382-5

JOURNAL OF MOLECULAR EVOLUTION

Molecular Phylogeny and Evolution of the Plant-Specific Seven-Transmembrane MLO Family

Alessandra Devoto,[1,*,†] H. Andreas Hartmann,[2,*] Pietro Piffanelli,[1] Candace Elliott,[1] Carl Simmons,[3] Graziana Taramino,[4] Chern-Sing Goh,[5] Fred E. Cohen,[5] Brent C. Emerson,[6] Paul Schulze-Lefert,[2] Ralph Panstruga[2]

[1] Sainsbury Laboratory, John Innes Centre, Colney, Norwich NR4 7UH, UK
[2] Max-Planck-Institut für Züchtungsforschung, Köln, Germany
[3] Pioneer, a DuPont Company, Johnston, IA 50131, USA
[4] E.I. DuPont de Nemours and Co., Wilmington, DE 19898, USA
[5] University of California, San Francisco, CA 94143, USA
[6] University of East Anglia, School of Biological Sciences, Norwich NR4 7TJ, UK

Received: 8 April 2002 / Accepted: 6 August 2002

Abstract. Homologues of barley *Mlo* encode the only family of seven-transmembrane (TM) proteins in plants. Their topology, subcellular localization, and sequence diversification are reminiscent of those of G-protein coupled receptors (GPCRs) from animals and fungi. We present a computational analysis of MLO family members based on 31 full-size and 3 partial sequences, which originate from several monocot species, the dicot *Arabidopsis thaliana*, and the moss *Ceratodon purpureus*. This enabled us to date the origin of the *Mlo* gene family back at least to the early stages of land plant evolution. The genomic organization of the corresponding genes supports a monophyletic origin of the *Mlo* gene family. Phylogenetic analysis revealed five clades, of which three contain both monocot and dicot members, while two indicate class-specific diversification. Analysis of the ratio of nonsynonymous-to-synonymous changes in coding sequences provided evidence for functional constraint on the evolution of the DNA sequences and purifying selection, which appears to be reduced in the first extracellular loop of 12 closely related orthologues. The 31 full-size sequences were examined for potential domain-specific intramolecular coevolution. This revealed evidence for concerted evolution of all three cytoplasmic domains with each other and the C-terminal cytoplasmic tail, suggesting interplay of all intracellular domains for MLO function.

Key words: Seven-transmembrane protein — Coevolution — Gene family — Exon/exon junctions — *Mlo* — G-protein coupled receptor

Introduction

In barley, presence of the wild-type *Mlo* gene modulates defense responses to the powdery mildew fungus, *Blumeria graminis* f. sp. *hordei* (Büschges et al. 1997). Homozygous *mlo* mutant plants exhibit full resistance to the fungal pathogen, whereas *Mlo* overexpression results in supersusceptibility (Wolter et al. 1993; Kim et al. 2002b). MLO is likely to have a role in additional biological processes since axenically

*These authors contributed equally to the work and are listed in alphabetical order,
Correspondence to: Ralph Panstruga, Max-Planck-Instutute für Züchtungsforschung, Department of Plant Microbe Interactions, Carl-von-Linné-Weg 10, 50829 Köln, Germany; e-mail: panstrug@mpiz-koeln.mpg.de
†*Present address:* University of East Anglia, School of Biological Sciences, Norwich NR4 7TJ UK

grown *mlo* mutant plants show accelerated leaf senescence symptoms and a spontaneous cell death phenotype (Wolter et al. 1993; Peterhänsel et al. 1997; Piffanelli et al. 2002). This suggests a function for MLO in cell death protection upon biotic stress and leaf senescence. Two genes, *Ror1* and *Ror2*, have been described that are required for full *mlo*-mediated resistance. Mutations in either of these genes confer partial susceptibility in an *mlo* mutant background and also compromise the spontaneous cell death phenotype (Freialdenhoven et al. 1996; Peterhänsel et al. 1997).

To date, MLO is the only plant polytopic membrane protein experimentally shown to consist of seven membrane-spanning domains (Devoto et al. 1999). However, a further protein, the putative GPCR GCR1, is predicted also to contain seven transmembrane (TM) helices (Josefsson and Rask 1997; Plakidou-Dymock et al. 1998). The barley MLO protein resides in the plasma membrane, with the N terminus positioned extracellularly and the C terminus intracellularly (Devoto et al. 1999). Database searches have revealed that MLO belongs to a gene family that is restricted to the plant kingdom. Inspection of the near-full-length *Arabidopsis* genome has shown that *Mlo*-like genes represent the only sequence-diversified family encoding seven-TM (7TM) proteins in plants, while *GCR1* is a single-copy gene (Devoto et al. 1999; The Arabidopsis Genome Initiative 2000). To date, all known animal and fungal (including yeast) sequence-diversified protein families with a 7TM topology function as G-protein coupled receptors (GPCRs), which relay extracellular signals into an intracellular response by activating a heterotrimeric G-protein (Bockaert and Pin 1999). Recent data, however, indicate that MLO-mediated defense suppression in barley functions independently of heterotrimeric G-proteins and that calmodulin interacts with MLO to dampen defense reactions against the powdery mildew fungus (Kim et al. 2002b).

Here we present a thorough computational analysis of the MLO protein family based on a comprehensive set of sequences derived from *Arabidopsis* and maize to trace back the phylogenetic history of these plant-specific proteins. We have investigated the data set for the presence of domain-specific adaptive molecular evolution. A recently developed algorithm that allows the identification of protein–protein interaction pairs identified candidate domains that have evolved in a concerted manner. Our findings are consistent with a presumptive receptor function of MLO proteins.

Materials and Methods

Mlo *DNA Sequences*

Mlo cDNA sequences from *Arabidopsis* were obtained by reverse transcriptase polymerase chain reaction (PCR) using oligonucleotides that were derived from the publicly available genomic sequences. Similarly, cDNAs of *TaMlo1*, *TaMlo2*, and *OsMlo2* and genomic sequences of *Mlo2* and *OsMlo1* were obtained using standard procedures (Elliott et al. 2002, further details about the isolation of these clones will be published elsewhere). Sequence information about Zea mays *Mlo* cDNAs (*ZmMlo1* to *-9*) were derived from corresponding expressed sequence tag (EST) clones from the combined DuPont/Pioneer EST collection. Nucleotide sequences of all cDNAs were determined by applying standard techniques on ABI373/377 automated sequencers.

Phylogenetic Analyses

Protein sequences were aligned using PileUp (Wisconsin Package Version 10.0; Genetics Computer Group, Madison, WI) and optimized by hand. Phylogenetic analyses were performed using the maximum parsimony search optimality criterion of PAUP* v.4.0b8 (Swofford 1998). Maximum parsimony analysis of protein sequences was performed for (i) full-length sequences excluding N and C termini, (ii) all transmembrane regions only, (iii) all extracellular and intracellular regions, (iv) all extracellular regions, and (v) all intracellular regions. An additional analysis was performed for a partial sequence alignment including an MLO homologue of a moss, *Ceratodon purpureus*. Searches were performed using the heuristic search option and all trees were rooted using the midpoint rooting option. Support for the branching arrangements was evaluated by bootstrap analyses using 1000 replicates.

Calculating d_N/d_S *Ratios*

To calculate the ratio of nonsynonymous-to-synonymous substitutions (d_N/d_S) we used the yn00 program of PAML (Yang 1997) implementing the method of Yang and Nielsen (2000). For these analyses we used an alignment of 1 wheat (*TaMlo2*) sequence and 11 sequences derived from nine species of the genus *Hordeum*. The *Hordeum* sequences correspond to amino acid residues 69–145 of barley MLO, covering the first extracellular loop and some neighboring residues, and were obtained by standard PCR amplification using genomic DNA as template and oligonucleotides Mlo4 5′-AAGGCGGAGCTCATGCTGGTGGGC-3′ and Mlo5 5′-ACGGCTTAGAGCTATGGTGATGAC-3′ as primers. Amplification products (~350–400 bp, including one intron) were purified on agarose gels, subcloned in pGEM-Teasy (Promega), and subjected to sequence analysis. We dissected the resulting nucleotide sequences (excluding primer and intron sequences) into three parts that were investigated separately: (i) the whole stretch, corresponding to amino acids 69–145 of barley MLO, (ii) extracellular loop 1 excluding the region between conserved cysteine residues 86 and 114, and (iii) the region between conserved cysteine residues 86 and 114. The yn00 program calculates d_N/d_S ratios for each pairwise comparison. We then summarized these as an average d_N/d_S ratio for each region (excluding ratios that had a zero value for either d_N or d_S) to compare differences in the rate of amino acid substitution among the three regions.

Coevolution Analysis

The correlation analysis was done on every possible domain–domain pair using methods described previously (Goh et al. 2000). Distance matrices were generated from the multiple alignments using ClustalW (Thompson et al. 1994). We employed a linear regression analysis measuring the correlation between pairwise evolutionary distances among all peptides in a multiple sequence alignment. These were correlated with the evolutionary distances among the corresponding binding partners using the linear correlation coefficient r (Pearson's correlation coefficient (Press et al.

Table 1. Compilation of *Mlo* homologues

Gene	Organism	GenBank accession No. cDNA	GenBank accession No. Genomic	Genome position	Introns	Amino acids
AtMlo1	*A. thaliana*	Z95352	At4g02600	Chr. IV, 15 cM	11	526
AtMlo2	*A. thaliana*	AF369563	At1g11310	Chr. I, 10 cM	13	573
AtMlo3	*A. thaliana*	AF369564	At3g45290	Chr. III, 61 cM	14	508
AtMlo4	*A. thaliana*	AF369565	At1g11000	Chr. I, 10 cM	14	570
AtMlo5	*A. thaliana*	AF369566	At2g33670	Chr. II, 76 cM	14	500
AtMlo6	*A. thaliana*	AF369567	At1g61560	Chr. I, 84 cM	13	583
AtMlo7	*A. thaliana*	AF369568	At2g17430	Chr. II, 32 cM	13	542
AtMlo8	*A. thaliana*	AF369569	At2g17480	Chr. II, 32 cM	14	593
AtMlo9	*A. thaliana*	AF369570	At1g42560	Chr. I, 62 cM	14	460
AtMlo10	*A. thaliana*	AF369571	At5g65970	Chr. V, 128 cM	14	569
AtMlo11	*A. thaliana*	AF369572	At5g53760	Chr. V, 100 cM	14	565
AtMlo12[a]	*A. thaliana*	AF369573	At2g39200	Chr. II, 72 cM	14	576
AtMlo13[b]	*A. thaliana*	AF369574	At4g24250	Chr. IV, 83 cM	13	478
AtMlo14	*A. thaliana*	AF369575	At1g26700	Chr. I, 38 cM	14	550
AtMlo15	*A. thaliana*	AF369576	At2g44110	Chr. II, 78 cM	13	496
CpMlo	*C. purpureus*	AW087034	—	n.d.	—	p.s.
Mlo	*H. vulgare*	Z83834	Y14573	Chr. IV	11	533
Mlo2	*H. vulgare*	—	Z95496	Chr. IV	11	544
OsMlo1	*O. sativa*	—	Z95353	Chr. VI	12	540
OsMlo2	*O. sativa*	AF384030	AP000615	Chr. III	12	555
OsMlo3	*O. sativa*	AF388195	—	n.d	—	554
OsMlo4	*O. sativa*	—	AC073166	Chr. X	14	580
TaMlo1	*T. aestivum*	AX063298	—	n.d.	—	534
TaMlo2	*T. aestivum*	AX063294	—	n.d	—	534
TaMlo3	*T. aestivum*	AX063296	—	n.d.	—	534
ZmMlo1	*Z. mays*	AY029312	—	Chr. I, bin 1	—	563
ZmMlo2	*Z. mays*	AY029313	—	Chr. I, bin 4	—	565
ZmMlo3	*Z. mays*	AY029314	—	Chr. II, bin 4	—	496
ZmMlo4	*Z. mays*	AY029315	—	Chr. III, bin 5	—	509
ZmMlo5	*Z. mays*	AY029316	—	Chr. III, bin 6	—	p.s.
ZmMlo6	*Z. mays*	AY029317	—	Chr. V, bin 4/5	—	515
ZmMlo7	*Z. mays*	AY029318	—	Chr. IX, bin 4	—	499
ZmMlo8	*Z. mays*	AY029319	—	Chr. VI, bin 5-7	—	492
ZmMlo9	*Z. mays*	AY029320	—	n.d.	—	p.s.

Note: p.s., partial sequence; —, genomic or cDNA sequence not available; n.d., not determined.
[a] Formerly designated *AtMlo18* (Devoto et al. 1999).
[b] Formerly designated *AtMlo20* (Devoto et al. 1999).

1998) between the distance matrices of all possible interacting domains where $-1 \leq r \leq +1$. Positive values of r indicate a positive correlation, and r values around zero indicate no correlation. Additionally, negative values of r indicate anticorrelation.

Results and Discussion

Phylogenetic Analysis of Mlo*-like Genes Suggests an Origin in the Early Stages of Land Plant Evolution*

Previously, we described the existence of *Mlo*-like sequences in different monocot and dicot species (Devoto et al. 1999). In the meantime, further genomic sequences and ESTs sequence-related to barley *Mlo* were released. By searching the public databases using the BLAST or PSI-BLAST algorithms (http://www.ncbi.nlm.nih.gov/BLAST/), *Mlo*-like genes were identified in an even broader range of monocotyledonous (*Hordeum vulgare*, *Oryza sativa*, *Secale cereale*, *Triticum aestivum*, *Zea mays*) as well as dicotyledonous plant species (*Arabidopsis thaliana*, *Brassica rapa*, *Citrullus lanatus*, *Glycine max*, *Gossypium hirsutum*, *Linum usatissimum*, *Lotus japonicus*, *Lycopersicon esculentum*, *Medicago truncatula*, *Solanum tuberosum*, *Sorghum bicolor*). Multiple distinct genes were found in most of these species, indicating their organization into multigene families.

Recently, the nearly full genomic sequence of *Arabidopsis thaliana* was released (The Arabidopsis Genome Initiative 2000), covering more than 90% of the 125-Mb genome of the weed. Based on this data, we identified 15 distinct members for which full-length genomic sequences are known (Table 1). The remaining 10 Mb of the Arabidopsis genomic sequence is supposed to cover mainly rDNA repeat units and centromeric and telomeric regions as well as other regions of complex sequence structure that are unlikely to harbor many coding sequences. Thus, we

conclude that the 15 Arabidopsis *Mlo* homologues identified to date are likely to represent the actual number. A former estimate of 25–35 homologues (Devoto et al. 1999) is apparently due to an overrepresentation of *Mlo* homologues in early released sequences of the *Arabidopsis* genome. The designation of the 15 genes is given in Table 1 (see also http://www.arabidopsis.org/info/genefamily/mlo.html). Only eight of these are currently represented by corresponding ESTs in GenBank, indicating their generally low expression levels. However, we were able to isolate matching cDNAs for all members by reverse transcriptase PCR. Subsequent DNA sequencing confirmed the identity of the clones, demonstrating that all 15 members are expressed, albeit at low levels (Table 1 and data not shown).

To identify *Mlo* family members in the monocotyledonous plant *Zea mays*, we searched the Pioneer/DuPont maize EST database, which to date comprises 400,000 ESTs. Nucleotide sequences of nine distinct *Mlo* genes were identified in this database (seven of which are full-length), indicative of a similar total number of *Mlo* genes in maize and *Arabidopsis*. Like *Arabidopsis*, most of the maize genes are expressed either at a low level or preferentially in particular tissues (data not shown).

Except for barley *Mlo*, no biological function has been assigned to any other *Mlo*-like gene to date. We have isolated cDNAs from wheat and rice that are exceptionally similar to barley *Mlo*. Due to their syntenic genomic locations relative to the barley gene on chromosome 4H, these members are likely to be orthologues (Elliott et al. 2002). In single-cell transfection experiments of barley *mlo* mutants (Shirasu et al. 1999), *OsMlo2* and *TaMlo2* showed either full (*TaMlo2*) or partial (*OsMlo2*) complementation, indicating that during evolution the function of these orthologues was preserved (Elliott et al. 2002). A comprehensive list of all 34 members analyzed here is shown in Table 1.

Phylogenetic analysis performed on 31 MLO full-length protein sequences identifies six subfamilies comprised of five clades (I–V), with strong bootstrap support for the monophyly of each clade, and a single divergent lineage (AtMLO3; Fig. 1). There is also strong bootstrap support for a sister group relationship between subfamily I and subfamily II, while relationships among the remaining subfamilies are unresolved. With a few exceptions, phylogenetic analyses of specific regions of the *Mlo* genes also recover these six subfamilies with moderate to high bootstrap support (Table 2). On average, subfamily members exhibit 45% identity and 70% similarity at the amino acid level. Interestingly, subfamily IV comprises only monocot homologues, including the presumptive orthologues from barley, wheat, and rice. Similarly, three *Arabidopsis* members (AtMLO2, AtMLO6, and AtMLO12) cluster together and define subfamily V, which appears to be restricted to dicots (or, alternatively, to *Arabidopsis*) given the fact that the analysis of 400,000 maize ESTs failed to reveal members of this gene cluster. The results of the phylogenetic analysis support an early evolutionary diversification of the MLO subgroups, well before the origin of monocots and dicots. MLO homologues of *Arabidopsis* and *Zea mays* are highly divergent, with representatives in clades I, II, III, and V and clades I, II, III, and IV, respectively. Maintenance of these subfamilies (clades) may indicate preservation of an early functional diversification. Whether monocot- and dicot-specific clades IV and V emerged after the separation of these two classes or whether members of these clades were lost subsequently remains elusive.

Since monocots are believed to have diverged from dicots approximately 100–270 million years ago (Wolfe et al. 1989; Schneider-Poetsch et al. 1998), *Mlo*-like genes must have already existed in their common progenitor. In fact, it would appear that this gene family is much older than the monocots and dicots. The monocot and dicot MLO sequences AtMLO4/ZmMLO4 and AtMLO1/ZmMLO8 group together as sister homologues with bootstrap values of 100 and 70, respectively (nodes A and B in Fig. 1). Unless these relationships are the result of horizontal gene transfer, the ages of these two nodes can be no younger than the 100 million- to 270 million-year divergence time between monocots and dicots. Several ESTs have been identified for the gymnosperm *Pinus taeda* demonstrating the presence of *Mlo* homologues in both subphyla of the spermatophyta (seed plants), angiosperms and gymnosperms, which are believed to have diverged from a common ancestor about 340–360 million years ago (Wolfe et al. 1989; Troitsky et al. 1991). Moreover, several ESTs (~20 of ~65,000) with a high sequence similarity to *Mlo* originate from the bryophyte *Physcomitrella patens*, and one (of ~1,700 ESTs) from the moss *Ceratodon purpureus*. A maximum parsimony analysis of an alignment based on the regions corresponding to the partial *C. purpureus* sequence (68 amino acids of the C terminus; Fig. 1) shows this sequence to fall within the diversity of monocots and dicots, with moderate bootstrap support for its placement within subfamily I. Bryophytes and tracheophytes (vascular plants) are believed to have diverged early in the evolution of green land plants, between the mid-Ordovician and the early Silurian period, approximately 400–450 million years ago (Wolfe et al. 1989; Kenrick et al. 1997). Thus, unless this is the result of horizontal gene transfer, a common ancestor of both must already have possessed an *Mlo* homologue and the node uniting CpMLO1/AtMLO4/ZmMLO4 (node C in Fig. 3) can be no younger than the 400 million- to 450 million-

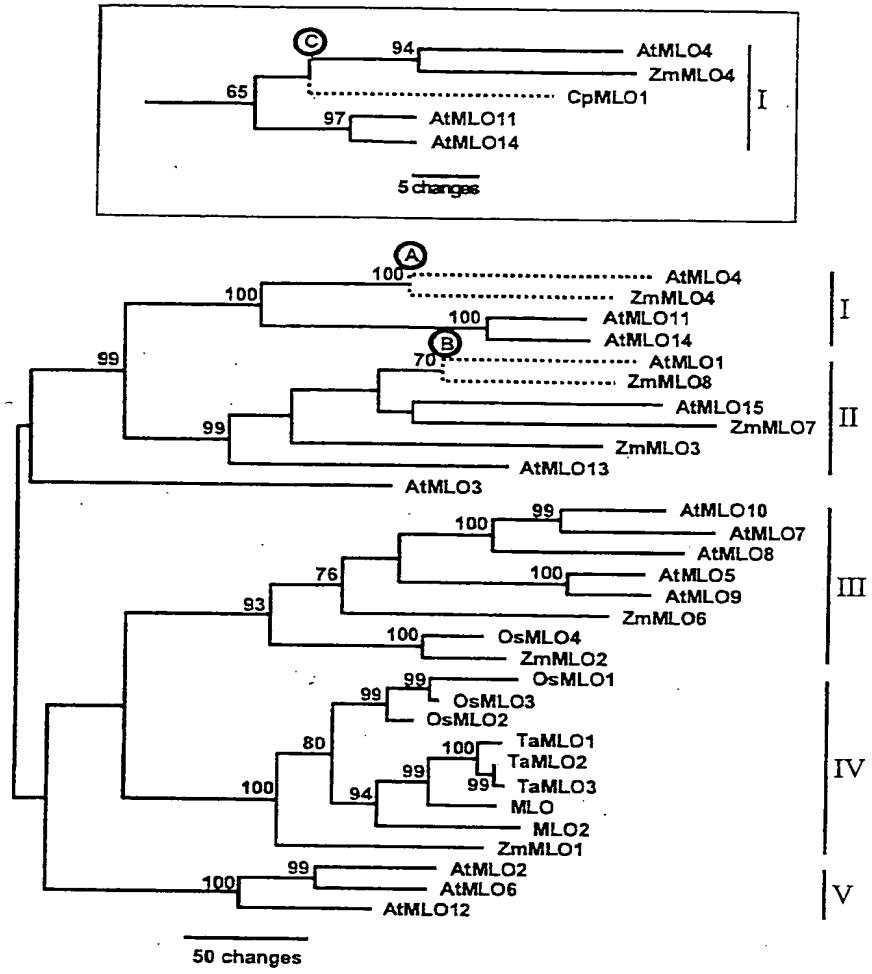

Fig. 1. Maximum parsimony phylogenetic analysis of amino acid sequence data for monocot and dicot MLO family members. Maximum parsimony tree constructed from full-length amino acid sequence data for MLO genes, excluding N and C termini. Branch lengths are proportional to the amount of amino acid changes. Numbers at the nodes indicate bootstrap support values (1000 replicates) above 60. Roman numerals denote major clades (subfamilies) referred to in the text. Nodes A and B indicate monocot and dicot sister lineages (*dashed lines*) referred to in the text. Inset: The phylogenetic position (node C) of the bryophyte MLO sequence (CpMLO1) from a maximum parsimony analysis of an alignment of partial sequences corresponding to the 68 amino acids of CpMLO1. The analysis included all MLO sequences in the partial alignment, but for clarity only clade I containing CpMLOl is shown.

Table 2. Bootstrap support values (1000 replicates) for monophyly of clades I–V from maximum parsimony analyses of specific regions of MLO protein sequences

MLO region analyzed	Bootstrap support value				
	Clade I	Clade II	Clade III	Clade IV	Clade V
Full protein excluding N and C termini	100	99	92	100	100
Intra- and extracellular regions	100	100	78	100	100
Transmembrane regions	87	73	59	99	100
Intracellular regions	99	90	66	95	100
Extracellular regions	95	55	< 50	99	100

year divergence time between bryophytes and tracheophytes.

We conclude from this observation that the presence of *Mlo* genes can be traced back at least to the early evolutionary stages of land plant development. This implies an ancient and vital function for the MLO family in plants. EST database searches (http://www.kazusa.or.jp/en/plant) of the unicellular green alga *Chlamydomonas reinhardtii* (37,990 ESTs) and the marine red alga *Porphyra yezoensis* (10,154 ESTs) detected no *Mlo*-like sequences in these two species. This could be the first evidence that *Mlo* emerged concurrently with the conquest of terrestrial habitats, although we cannot rule out the possibility that the number of currently available algal ESTs is too low to identify *Mlo*-like sequences.

Closely related members belonging to the same subfamily but originating from different species may be identified as orthologues with similar functions, as demonstrated experimentally for MLO, TaMLO2, and OsMLO2 (see above; Elliott et al. 2002). Whether the observed clustering correlates generally with a common function of the members is currently under investigation.

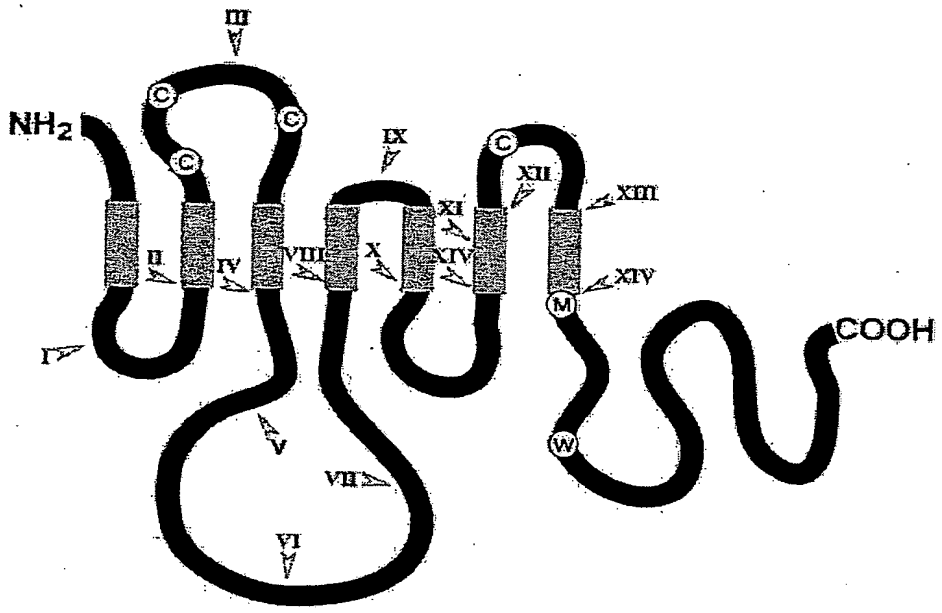

Fig. 2. Scheme of the MLOs proteins. *Gray boxes* designate the seven TM helices. *Arrowheads* indicate the position of splice junctions (exon/exon junctions at the protein level), with the corresponding introns numbered with Roman numerals. C, M, and W denote conserved cysteine, methionine, and tryptophan residues, respectively.

A Common Scaffold Topology Accommodates Two Hypervariable Domains

A hallmark of all MLO family members is the presence of seven TM domains. The predictions obtained for each of the full-size family members from Table 1 using the TMHMM algorithm (Sonnhammer et al. 1998) exactly matched the 7TM topology determined experimentally for the barley MLO protein (Devoto et al. 1999). Similarly, the predicted distribution of the amino acid residues with respect to the membrane is comparable to that for the barley protein: generally 50–60% of the protein is predicted to be cytoplasmic, 20–30% to be embedded in the membrane, and the rest is thought to be extracellular/lumenal. These observations indicate a shared scaffold topology for all MLO protein family members, consisting of seven TM helices, an N-terminal extracellular or lumenal end, three cytoplasmic and three extracellular/lumenal loops, and a cytoplasmic C-terminal tail (Fig. 2). Although a rice MLO homologue has also been shown to reside within the plasma membrane (Kim et al. 2002a), the scaffold topology does not provide conclusive evidence for a common subcellular localization. For simplicity, we refer in the following to "extracellular" rather than to "extracellular/lumenal" domains.

Another characteristic is the presence of four strictly conserved cysteine residues in extracellular loops 1 and 3 (Fig. 2). If these cysteine residues form (a) disulfide bridge(s) either with each other or with the two other invariant extracellular cysteines, this domain could subsequently form an exposed loop/ligand binding site. This is frequently found in mammalian 7TM receptors to stabilize the relative arrangement of the TM helices to each other (Probst et al. 1992; Strader et al. 1994). Extraordinary length variability occurs between cysteine residues 99 and 115 in extracellular loop 1, contributing to an exceptional sequence variation in this region among family members (Fig. 3A). The C terminus defines the second domain, which is highly variable in both sequence and length (ranging from 55 to 253 amino acid residues; Fig. 3B). However, the first ~25 residues proximal to TM VII are rather conserved, harboring the recently discovered calmodulin binding site present in MLO proteins (Fig. 2) (Kim et al. 2002a, b). A hallmark of this binding site is a strictly conserved tryptophan residue that has been demonstrated to be essential for the interaction with calmodulin (Figs. 2 and 3B) (Kim et al. 2002a, b).

Sequence Diversity in Extracellular Loop 1 and Reduced Functional Constraint

The comparatively high level of sequence variability observed in extracellular loop 1 can be interpreted in two ways: either this region determines specificity of individual MLO members by creating unique binding sites for putative ligands or this region has no isoform-specific function but serves as a structural component of the 7TM family. In the latter case, the observed sequence variability would be the result of evolution by random drift, while in the former, it would reflect selection toward isoform specificity. To distinguish between these alternatives, the ratio d_N/d_S of nonsynonymous (amino acid-changing; d_N)-to-synonymous (silent; d_S) substitutions per nonsynonymous and synonymous sites is a suitable indicator. Pseudogenes without any evolutionary selective pressure will accumulate neutral and amino acid-changing substitutions in their DNA sequence at the same frequency,

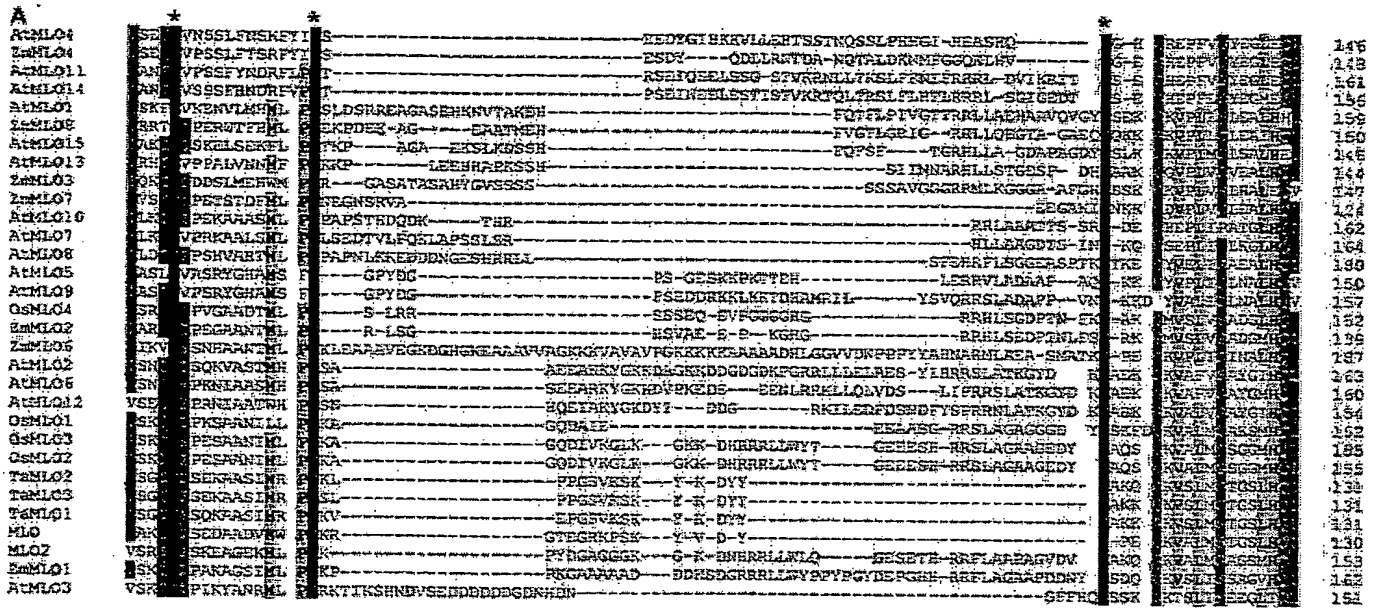

Fig. 3. Multiple sequence alignment of MLO proteins. **A** Alignment of amino acid sequences corresponding to the first extracellular loop. **B** Alignment of amino acid sequences corresponding to TM VII and the C-terminal tail. Sequences were aligned using PileUp (Winsconsin Package Version 10.0; Genetics Computer Group, Madison, WI); spaces were manually introduced to increase the similarity in the alignment. *Shading* indicates the degree of conservation between amino acids. Identical amino acid residues (100% conserved) are shaded in *black*; 80% or greater conserved, 60% or greater conserved, and less than 60% conserved residues are shaded in *dark gray*, *light gray*, and *white*, respectively. Numbers indicate amino acid positions within the protein; *asterisks* indicate conserved cysteine residues; *black* and *gray triangles* indicate the methionine corresponding to the start of the last exon and the conserved tryptophan residue of the calmodulin binding domain, respectively.

resulting in a d_N/d_S ratio of ~1. In contrast, in the majority of genes most of the occurring nonsynonymous changes are probably deleterious, resulting in purifying counter-selection. In these cases, synonymous substitutions take place more often than nonsynonymous ones, resulting in a dN/dS ratio of < 1. As a third possibility, certain coding regions are selected for extraordinary high rates of nonsynonymous substitutions (resulting in a d_N/d_S ratio of > 1). This behavior is true for fast-evolving genes that underlie adaptive molecular evolution as, for example, several surface antigens of pathogens and the matching defense systems in the corresponding hosts (Yang and Bielawski 2000). Since this method provides reliable results only if the sequences investigated are neither too similar nor too different (Yang and Bielawski 2000), we first had to select suitable sequences. Known full-size MLO sequences were unsuitable because they are highly divergent in extracellular loop 1 (Fig. 3A). We PCR-amplified a fragment of the *Mlo* genomic sequence (corresponding to extracellular loop 1 and some flanking amino acid residues) from eight species of the genus *Hordeum* (Materials and Methods; Table 3 and Fig. 4). In two cases, we obtained two distinct sequences each, likely reflecting the polyploid nature of these species. The resulting predicted amino acid sequences are only moderately divergent in extracellular loop 1 and thus ideally suited for d_N/d_S analyses (compare Figs. 3A and 4).

We calculated d_N/d_S ratios for each possible pair of 10 amplified and 2 previously known sequences of (i) the region corresponding to amino acid residues 69–145 of barley MLO, (ii) extracellular loop 1 excluding the region between conserved cysteines 86 and 114, and (iii) the region between conserved cysteines 86 and 114. The average d_N/d_S ratio values for each of these are 0.138, $\sigma = 0.048$ (i), 0.154, $\sigma = 0.054$ (ii), and 0.275, $\sigma = 0.170$ (iii). All of these values are well below 1, indicating functional constraint on the evolution of the DNA sequences and purifying selection. However, it appears that functional constraint is less for the region between conserved cysteine residues 86 and 114 in extracellular loop 1, as the average d_N/d_S ratio for this section is almost two times higher than that of its 5′ and 3′ flanking sequences (0.275 versus 0.154, respectively), although this difference is not statistically significant. This result can be interpreted in two ways. It may indicate that relaxed constraint in DNA evolution causes, over long periods of time, the sequence variation found among compiled MLO family members. Alternatively, in this particular case, the d_N/d_S ratio might not be a reliable indicator for the molecular mechanism leading to the observed variability. It will be interesting to find out whether differences in this region correspond to the ability to bind diverse interacting partners in the extracellular space.

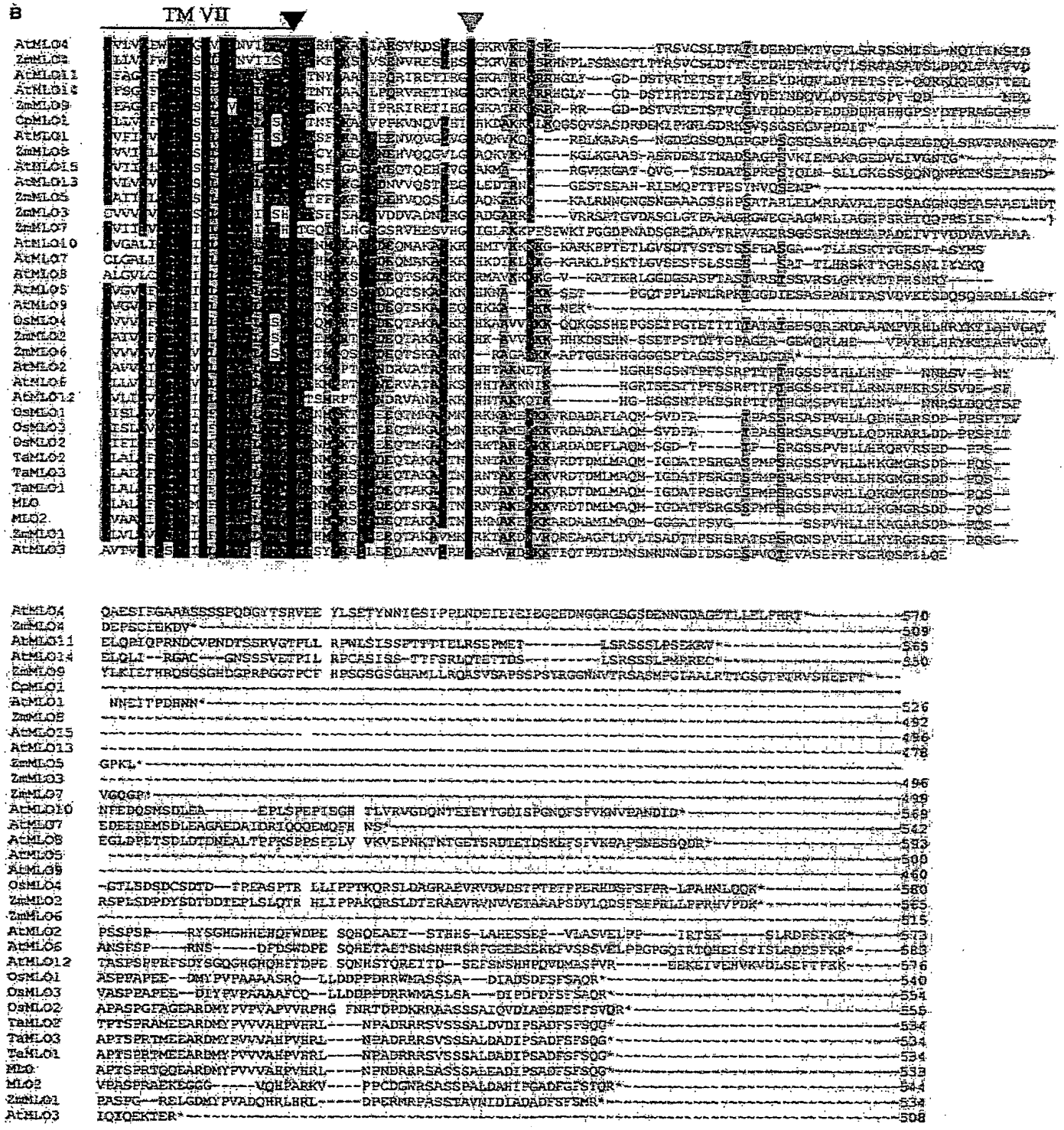

Fig. 3. (Continued)

Structural Organization of Mlo *Genomic Sequences Provides Further Evidence for a Monophyletic Origin of the Gene Family*

A comparison of the gene structure among available full genomic sequences of family members revealed 11 to 14 introns per *Mlo* gene (Table 1). Most of the introns are 80 to 90 nucleotides in size, with no sequence conservation even among phylogenetically closely related members. It is noticeable that in all but one case the exon/exon junctions map exactly at the identical position at the corresponding protein level, supporting a monophyletic origin for the gene family (Fig. 2). The only exception is represented by intron

Table 3. Sequences of *Hordeum* species used for the d_N/d_S analysis

Species	Ploidy[a]	GenBank accession No.(s.)
H. vulgare	Diploid	Z83834
H. vulgare f. *agriocrithon*	Diploid	AY090646
H. vulgare ssp. *spontaneum*	Diploid	AY090647
H. brevisbulatum	Diploid, tetraploid, and hexaploid	AY090638, AY090639
H. bulbosum	Diploid and tetraploid	AY090641, AY090642
H. chilense	Diploid	AY090643
H. jubatum	Tetraploid	AY090640
H. murinum ssp. *murinum*	Tetraploid	AY090645
H. murinum ssp. *leporinum*	Tetraploid and hexaploid	AY090644

[a] According to von Bothmer et al. (1995).

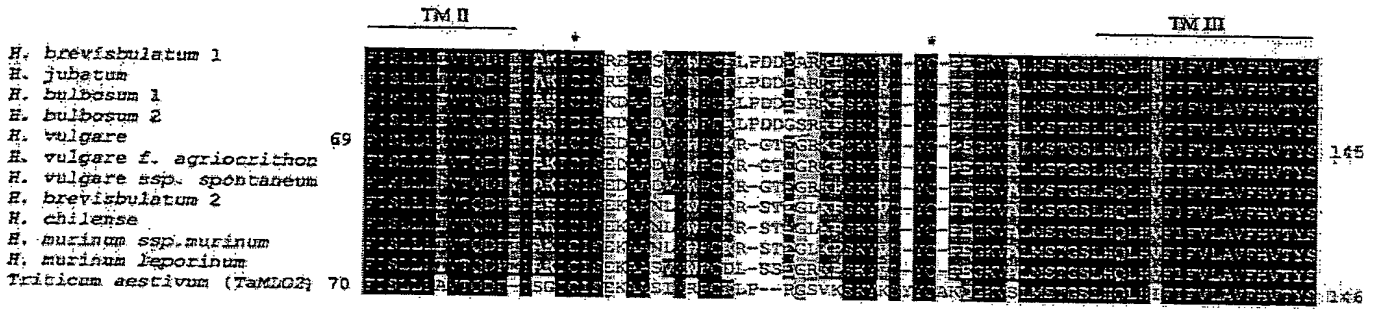

Fig. 4. Amino acid sequence alignment of MLO sequences used for the d_N/d_S analysis. Amino acid sequences corresponding to extracellular loop 1 and flanking regions from 11 presumptive orthologues of nine species of the genus *Hordeum* and a wheat sequence were aligned using ClustalW. Identical amino acid residues (100% conserved) are shaded in *black*; 80% or greater conserved, 60% or greater conserved, and less than 60% conserved residues are shaded in dark *gray*, *light gray*, and *white*, respectively. The two *asterisks* indicate conserved cysteines that are at positions 86 and 114 in barley MLO.

V, which is located at a slightly different position in *AtMlo1*, *-13*, and *-15*. Intron VI is absent in *AtMlo2* and *AtMlo6*, while intron XI is missing in *AtMlo1*, *-13*, and *-15*. These observations are in full agreement with the phylogenetic analysis (see above and Fig. 1), suggesting that highly similar members within *Arabidopsis* did not arise by convergence from different progenitor sequences but diverged from a single common ancestor gene. The C-terminal tails are always encoded by a single exon, invariably starting with a consensus translational initiation sequence including the start codon ATG (Figs. 2 and 3B). Whether this reflects an ancient gene shuffling event remains speculative.

The splice junctions in the gene family map mainly to the boundaries between the encoded loop and transmembrane regions (Fig. 2). Eight of the 14 exon/exon junctions are located proximal or distal to the transmembrane helical termini. Only one TM helix (VI) is interrupted by a splice junction. The remaining junctions are located within extracellular loop 1, intra- and extracellular loop 2, and TM helix VI. No exon–exon junction was observed in the amino- and the carboxyl-terminal ends of the family members proximal to the first TM helix or distal to TM VII. The fact that individual TM helices are encoded by single exons is common to other polytopic membrane proteins (Argos and Rao 1985; Miao and Verma 1993). This is thought to reflect their role as an evolutionary unit that is subject to severe selection constraint to maintain the structurally stable, multihelical TM core. Such a unit may serve as a module to create variability in the number of TM helices of polytopic membrane proteins (e.g., by exon shuffling).

AtMlo *Distribution in the* Arabidopsis *Genome*

It has been demonstrated recently that most of the genome of *Arabidopsis thaliana* is internally duplicated, indicating *Arabidopsis* as a potential ancient tetraploid species (Blanc et al. 2000; The Arabidopsis Genome Initiative 2000; Vision et al. 2000). Additionally, Vision et al. (2000) provided evidence that the current state of the *Arabidopsis* genome may result from at least four large-scale duplication events that took place 100 to 200 million years ago. These duplication processes must have also involved chromosome fusions resulting in extended genomic regions in which the number, order, and orientation of duplicated genes are preserved. After duplication, affected regions were subject to extensive subchromosomal rearrangements, such as inversions, translocations, and loss or transposition of single genes or groups of genes.

We investigated the distribution of *AtMlo* genes in extended duplicated genomic regions to identify pu-

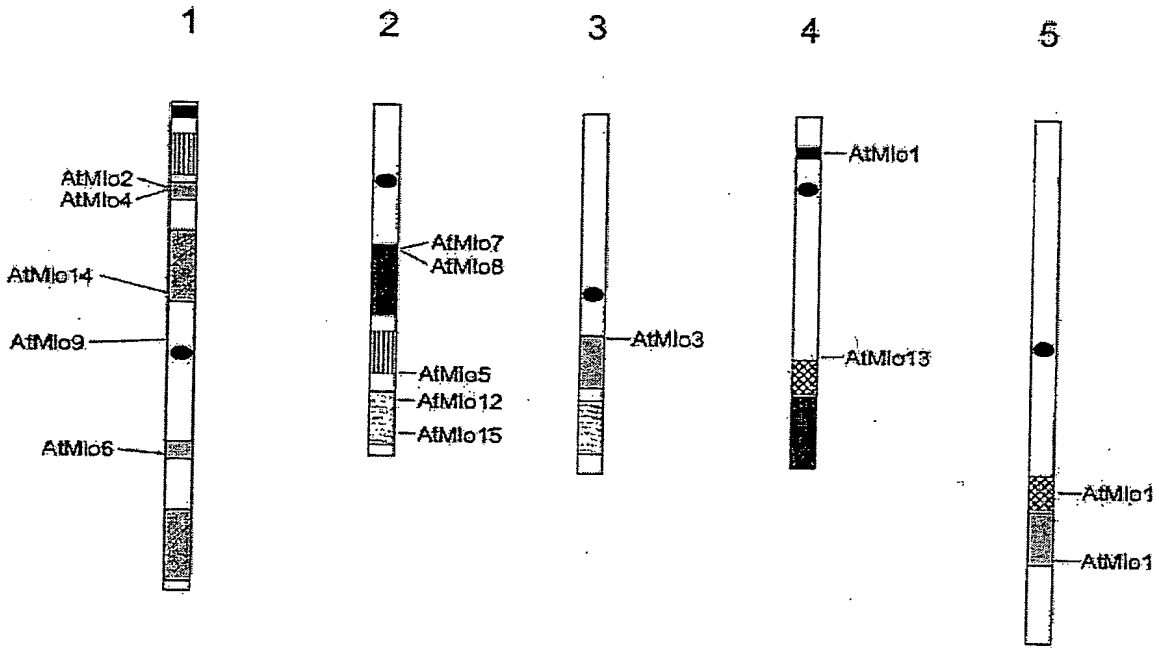

Fig. 5. Distribution of *AtMlo* members in the *Arabidopsis* genome. The five chromosomes of *Arabidopsis thaliana* are schematically represented by *rectangles* numbered from 1–5. Centromeric regions are indicated by *black ovals*. *Marked blocks* indicate areas of large-scale genome duplications. Relative positions of the 15 *AtMlo* genes are shown. (Adapted from Blanc et al. 2000).

tative functionally redundant copies of *AtMlo* genes. For this analysis we used the template map of *Arabidopsis* genomic duplications described by Blanc et al. (2000) because the start and end of the copied regions are exactly designated by particular BAC clones. We found that *Mlo* genes are located on all five chromosomes without any obvious clustering. With two exceptions (*AtMlo9* and *-13*), all *AtMlo* genes are located within regions that are supposed to have undergone a previous large-scale duplication event (Fig. 5). Unexpectedly, *AtMlo* genes were always found as a single copy in the duplicated areas, except *AtMlo2* and *AtMlo6*, for which the number, order, and orientation of flanking genes are rather conserved. Although it is known that less than half of the genes (37–47%, depending on the significance criteria) in the duplicated areas are conserved in their corresponding copy region (Bancroft 2001), *AtMlo* genes behave differently because of only a single recognizable duplication. Whether this indicates constraints in copy numbers or exceptionally high micro-translocation/deletion events cannot be resolved. Taken together, this approach identifies only AtMLO2 and AtMLO6 as the result of a large-scale duplication event. It should be interesting to find out whether these two genes are functionally redundant or whether the few sequence differences lead to functional diversification.

Coevolution Among Domains of MLO Proteins

Recently, Goh et al. (2000) have developed an algorithm that allows the identification of protein–protein interaction pairs and can be adapted to the assessment of intramolecular coevolution of peptide domains within a single protein family. The method is based on the assumption that if there are two domains within a single protein that have to act cooperatively for proper function, evolutionary changes of the amino acid sequence within one of the domains will either result in counter-selection or in compensating changes in the amino acid sequence of the other domain. In terms of evolution, these two domains will evolve in a coordinated manner, resulting in a linked phylogenetic relationship. If there is no cooperation between the two domains, they are believed to evolve independently resulting in an unlinked phylogenetic relationship. The algorithm has been used by Pazos and Valencia (2001) to test the impact of the method by analyzing potential intramolecular interactions of structural domains in bipartite proteins and by investigating known protein–protein interaction pairs. The authors conclude from their results that the procedure is capable of detecting true interactions in $> 66\%$ of the cases if a correlation > 0.8 is detected.

We dissected 31 full-length sequences of MLO proteins into their single peptide domains. This procedure resulted in 15 sets of peptide sequences, representing the N and C termini, the seven TM regions, the three cytoplasmic, and the three extracellular loops. We paired each set of peptide sequences with each other and calculated correlation coefficients for all 105 possible pairings (Fig. 6 and Materials and Methods). The observed correlation coefficients were

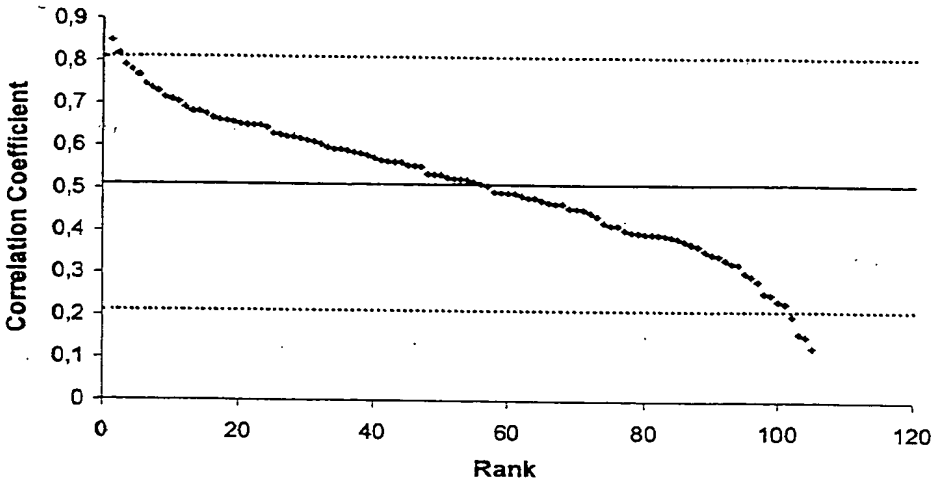

Fig. 6. Interdomain correlation analysis of MLO proteins. Correlation coefficients of all 105 interdomain pairings of the 15 sets of peptide domains from 31 MLO proteins were plotted against the relative ranking (ranging from 1 to 105) of the respective pair. Mean values and 1.96× standard deviations are indicated by the *bold horizontal line* and *dotted horizontal lines*, respectively.

in the range of 0.15 to 0.85, with an average of 0.51 and a standard deviation of 0.15 (Fig. 6). In Table 4 we have listed the top five pairings with the highest correlation coefficients, which we will discuss in detail. All of them have values close to or even above the 1.96 times standard deviation boundary (marking significant values with a probability value of $p < 0.05$). This indicates that coevolution between the respective peptide domains is likely. Among these top five pairs, the three possible combinations among the cytoplasmic domains IC2 and IC3 and the C terminus have the highest scores (Table 4), about 0.8, a value that has been suggested to be a good empirical cutoff to indicate with a high probability true positive interactions (Pazos and Valencia, 2001). The following two pairs both indicate also a possible coevolution of IC1 with loops IC2 and IC3 (Table 4). Taken together, the analysis provides evidence for coevolution of all cytoplasmic loops with the C terminus, showing a particular emphasis on IC2, IC3, and the C terminus. Probable coevolution between the cytoplasmic domains of MLO suggests interplay of these domains and interaction with a putative partner(s) for MLO protein function. Although other scenarios are possible, the most likely interpretation is related to a conserved interaction of the cytoplasmic domains with a common binding partner. An analogous situation has been demonstrated experimentally for the well-characterized family of GPCRs in binding heterotrimeric G-proteins (reviewed by Hamm 1998). The relative absence of correlations joining the extracellular domains could relate to the heterogeneity of presumptive ligands that might bind and activate MLO proteins.

Table 4. Correlation coefficients of the coevolution analysis of MLO protein domains

Rank	Pair	Correlation coefficient
1	IC3/C terminus	0.85
2	IC2/IC3	0.82
3	IC2/C terminus	0.79
4	IC1/IC2	0.78
5	IC1/IC3	0.77

Note. IC, intracellular loop.

GenBank Accession Numbers

GenBank accession numbers for newly deposited sequences are as follows: Z95353 (*OsMlo1*), AF384030 (*OsMlo2*), AF361933 (*TaMlo1*), AF361932 (*TaMlo2*), Z95352 and AF369563–AF369576 (*AtMlo1–AtMlo15*), AY029312–AY029320 (*ZmMlo1–ZmMlo9*), and AY090638–AY090647 (*Hordeum* species in Table 3).

Acknowledgments. We thank our colleague Günter Theissen for helpful and critical comments on the manuscript. We are grateful to Brigitte Schauf (Max-Planck Institut für Züchtungsforschung), Robin Wineland (Pioneer), and Wayne Powell (DuPont) for their assistance. This work was supported by grants from the Gatsby Charitable Foundation, the BBSRC, and the Max-Planck Society to P.S.-L. A.D. was supported by an EU fellowship (Training and Mobility of Researcher Program in Biotechnology); H.A.H., by an EMBO fellowship. C.-S.G. and F.E.C. were supported by Grant FEC GM39900 from the National Institutes of Health.

References

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Argos P, Rao JKM (1985) Relationships between exons and the predicted structure of membrane-bound proteins. Biochim Biophys Acta 827:283–297

Bancroft I (2001) Duplicate and diverge: The evolution of plant genome microstructure. Trends Genet 17:89–93

Blanc G, Barakat A, Guyot R, Cooke R, Delseny I (2000) Extensive duplication and reshuffling in the Arabidopsis genome. Plant Cell 12:1093–1101

Bockaert J, Pin JP (1999) Molecular tinkering of G protein-coupled receptors: An evolutionary success. EMBO J 18:1723–1729

Büschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, van Daelen R, van der Lee T, Diergaarde P, Groenendijk J, Töpsch S, Vos P, Salamini F, Schulze-Lefert P (1997) The barley *Mlo* gene: A novel control element of plant pathogen resistance. Cell 88:695–705

Devoto A, Piffanelli P, Nilsson I, Wallin E, Panstruga R, von Heijne G, Schulze-Lefert P (1999) Topology, subcellular localization, and sequence diversity of the *Mlo* family in plants. J Biol Chem 274:34993–35004

Elliott C, Zhou F, Spielmeyer W, Panstruga R, Schulze-Lefert P (2002) Functional conservation of wheat and rice Mlo orthologs in defence modulation to the powdery mildew fungus. Mol Plant-Microbe Interact 15:1069–1077

Freialdenhoven A, Peterhänsel C, Kurth J, Kreuzaler F, Schulze-Lefert P (1996) Identification of genes required for the function of non-race-specific *mlo* resistance to powdery mildew in barley. Plant Cell 8:5–14

Goh CS, Bogan AA, Joachimiak M, Walther D, Cohen FE (2000) Co-evolution of proteins with their interaction partners. J Mol Biol 299:283–293

Hamm HE (1998) The many faces of G protein signaling. J Biol Chem 273:669–672

Josefsson LG, Rask L (1997) Cloning of a putative G-protein-coupled receptor from *Arabidopsis thaliana*. Eur J Biochem 249:415–420

Kenrick P, Crane PR (1997) The origin and early evolution of plants on land. Nature 389:33–39

Kim MC, Lee SH, Kim JK, Chun HJ, Ok HM, Moon BC, Kang CH, Chung WS, Park CY, Choi MS, Kang YH, Koo SC, Koo YC, Jung JC, Schulze-Lefert P, Cho MJ (2002a) *Mlo*, a modulator of plant defense and cell death, is a novel calmodulin-binding protein: Isolation and characterization of a rice *Mlo* homologue. J Biol Chem 277:19304–19314

Kim MC, Panstruga R, Elliott C, Müller J, Devoto A, Yoon HW, Park HC, Cho MJ, Schulze-Lefert P (2002b) Calmodulin interacts with MLO protein to regulate defence against mildew in barley. Nature 416:447–450

Miao GH, Verma DPS (1993) Soybean nodulin-26 gene encoding a channel protein is expressed only in the infected cells of nodules and is regulated differently in roots of homologous and heterologous plants. Plant Cell 5:781–794

Pazos F, Valencia A (2001) Similarity of phylogenetic trees as indicator of protein-protein interaction. Prot Eng 14:609–614

Peterhänsel C, Freialdenhoven A, Kurth J, Kolsch R, Schulze-Lefert P (1997) Interaction analyses of genes required for resistance responses to powdery mildew in barley reveal distinct pathways leading to leaf cell death. Plant Cell 9:1397–1409

Piffanelli P, Zhou F, Casais C, Orme J, Schaffrath U, Collins N, Panstruga R, Schulze-Lefert P (2002) The barley MLO modulator of defence and cell death is responsive to biotic and abiotic stress stimuli. Plant Physio 129:1076–1085

Plakidou-Dymock S, Dymock D, Hooley R (1998) A higher plant seven-transmembrane receptor that influences sensitivity to cytokinins. Curr Biol 8:315–324

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1998) Numerical recipes in C. Cambridge University Press, Cambridge, UK

Probst WC, Snyder LA, Schuser DI, Brosius J, Sealfon SC (1992) Sequence alignment of the G-protein coupled receptor superfamily. DNA Cell Biol 11:1–20

Schneider-Poetsch HAW, Kolukisaoglu U, Clapham DH, Hughes J, Lamparter T (1998) Nonangiosperm phytochromes and the evolution of vascular plants. Physiol Plant 102:612–622

Shirasu K, Nielsen K, Piffanelli P, Oliver R, Schulze-Lefert P (1999) Cell-autonomous complementation of *mlo* resistance using a biolistic transient expression system. Plant J 17:293–299

Sonnhammer ELL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. In: Glasgow J, Lathorp R, Littlejohn T, Major F (eds) Proceedings, Sixth International Conference on Intelligent Systems for Molecular Biology. AAAI Press, Menlo Park, CA, pp 175–182

Strader CD, Fong TM, Tota MR, Underwood D, Dixon RAF (1994) Structure and function of G-protein-coupled receptors. Annu Rev Biochem 63:101–132

Swofford DL (1998) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Sunderland, MA

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal-W—Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Troitsky AV, Melekhovets YF, Rakhimova GM, Bobrova VK, Valiejoroman KM, Antonov AS (1991) Angiosperm origin and early stages of seed plant evolution deduced from rRNA sequence comparisons. J Mol Evol 32:253–261

Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in Arabidopsis. Science 290:2114–2117

von Bothmer R, Jacobsen N, Baden C, Jøgensen RB, Linde-Laursen I (1995) An ecogeographical study of the genus *Hordeum*. Systematic and ecogeographic studies on crop genepools 7, 2nd ed. International Plant Genetic Resources Institute, Rome

Wolfe KH, Gouy ML, Yang YW, Sharp PM, Li WH (1989) Date of the monocot dicot divergence estimated from chloroplast DNA-sequence data. Proc Natl Acad Sci USA 86:6201–6205

Wolter M, Hollricher K Salamini F, Schulze-Lefert P (1993) The *mlo* resistance alleles to powdery mildew infection in barley trigger a developmentally controlled defense mimic phenotype. Mol Gen Genet 239:122–128

Yang ZH (1997) PAML: A program package for phylogenetic analysis by maximum likelihood. CABIOS 13:555–556

Yang ZH, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. Trends Ecol Evol 15:496–503

Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17:32–43

Appendix A

Appendix A shows a comparison of the amino acid sequences of Mlo proteins SEQ ID NO:32 and barley (NCBI General Identifier No. 1877221 , SEQ ID NO: 39). Amino acids conserved among both sequences are indicated with an asterisk (*) on the top row; dashes are used by the program to maximize alignment of the sequences. The seven membrane-spanning helices (underlined), the conserved Cystein residues (▲) the putative nuclear localization motif (bold letters) and the two casein kinase II motifs (boxed sequences) were all identified and discussed by Bueschges et al. (Cell 1997, 88: 695-705) and Devoto et al. (JBC 1999, 274: 34993-35004). Sites of mutations as published in Bueschges et al. (Cell 1997, 88: 695-705) are indicated in alignment as follows: (●) mutations, (►) frameshift changes, (Δ) deletions.

```
                *       *** ********** *** *  **** ** ********  *** ** *****
SEQ ID NO:32    MAEDYEYPPARTLPETPSWAVALVFAVMIIVSVLLEHALHKLGHWFHKRHKNALAEALEK
Gi:1877221      MSDKKG-VPARELPETPSWAVAVVFAAMVLVSVLMEHGLHKLGHWFQHRHKKALWEALEK
                ●       ●                ●●
                 ************** ******   ***** **  * **       * *** ** *   *
SEQ ID NO:32    IKAELMLVGFISLLLAVTQDPISG-ICISEKAASIMRPCSLPPGSVK-SKYKDYYCAKKG
Gi:1877221      MKAELMLVGFISLLLIVTQDPIIAKICISEDAADVMWPCKRGTEGRKPSKYVDY-CPE-G
                                        ▲          ▲              ▲
                ** *********** ***************  ***************** **********
SEQ ID NO:32    KVSLMSTGSLHQLHIFIFVLAVFHVTYSVIIMALSRLKMRTWKKWETETASLEYQFANDP
Gi:1877221      KVALMSTGSLHQLHVFIFVLAVFHVTYSVITIALSRLKMRTWKKWETETTSLEYQFANDP
                                                    ►  ●
                ********************************************************* ***
SEQ ID NO:32    ARFRFTHQTSFVKRHLGLSSTPGIRWVVAFFRQFFRSVTKVDYLTLRAGFINAHLSHNSK
Gi:1877221      ARFRFTHQTSFVKRHLGLSSTPGIRWVVAFFRQFFRSVTKVDYLTLRAGFINAHLSQNSK
                    ΔΔ                                          ●
                ************************************ * *** ****************
SEQ ID NO:32    FDFHKYIKRSMEDDFKVVVGISLPLWCVAILTLFLDIDGIGTLTWISFIPLVILLCVGTK
Gi:1877221      FDFHKYIKRSMEDDFKVVVGISLPLWGVAILTLFLDINGVGTLIWISFIPLVILLCVGTK
                                          ●
                ************************************************************
SEQ ID NO:32    LEMIIMEMALEIQDRASVIKGAPVVEPSNKFFWFHRPDWVLFFIHLTLFQNAFQMAHFVW
Gi:1877221      LEMIIMEMALEIQDRASVIKGAPVVEPSNKFFWFHRPDWVLFFIHLTLFQNAFQMAHFVW

                ********** *  ********* ********** ************************
SEQ ID NO:32    TVATPGLKKCFHMHIGLSIMKVVLGLALQFLCSYITFPLYALVTQMGSNMKRSIFDEQTA
Gi:1877221      TVATPGLKKCYHTQIGLSIMKVVVGLALQFLCSYMTFPLYALVTQMGSNMKRSIFDEQTS
                          ▲
                ********************************** ****** *****************
SEQ ID NO:32    KALTNWRNTAKEKKKVRDTDMLMAQMIGDATPSRGTSPMPSRASSPVHLLHKGMGRSDDP
Gi:1877221      KALTNWRNTAKEKKKVRDTDMLMAQMIGDATPSRGSSPMPSRGSSPVHLLHKGMGRSDDP

                *********  ******************* ***** ***** *************
SEQ ID NO:32    QSAPTSPRTMEEARDMYPVVVAHPVHRLNPADRRRSVSSSALDADIPSADFSFSQG
Gi:1877221      QSAPTSPRTQQEARDMYPVVVAHPVHRLNPNDRRRSASSSALEADIPSADFSFSQG
```